SpeDo: 6 DOF Ego-Motion Sensor Using Speckle Defocus Imaging

Kensei Jo
Columbia University
kenseij@cs.columbia.edu

Mohit Gupta
Columbia University
mohitg@cs.columbia.edu

Shree K. Nayar
Columbia University
nayar@cs.columbia.edu

Abstract

Sensors that measure their motion with respect to the surrounding environment (ego-motion sensors) can be broadly classified into two categories. First is inertial sensors such as accelerometers. In order to estimate position and velocity, these sensors integrate the measured acceleration, which often results in accumulation of large errors over time. Second, camera-based approaches such as SLAM that can measure position directly, but their performance depends on the surrounding scenes properties. These approaches cannot function reliably if the scene has low frequency textures or small depth variations. We present a novel ego-motion sensor called SpeDo that addresses these fundamental limitations. SpeDo is based on using coherent light sources and cameras with large defocus. Coherent light, on interacting with a scene, creates a high frequency interferometric pattern in the captured images, called speckle. We develop a theoretical model for speckle flow (motion of speckle as a function of sensor motion), and show that it is quasi-invariant to surrounding scenes properties. As a result, SpeDo can measure ego-motion (not derivative of motion) simply by estimating optical flow at a few image locations. We have built a low-cost and compact hardware prototype of SpeDo and demonstrated high precision 6 DOF ego-motion estimation for complex trajectories in scenarios where the scene properties are challenging (e.g., repeating or no texture) as well as unknown.

1. Introduction

Measuring the motion of an object relative to the surrounding world has several applications, such as robot navigation (e.g., self-driving cars and autonomous drones) and user-interface (e.g., optical mouse and augmented reality displays). Ego-motion sensors (or odometers) are self-contained motion sensors that can be attached to the target object itself and thus can measure object motion without requiring any external devices. Current ego-motion sensors can be broadly classified into two categories. The first class is sensors based on inertia called inertial measurement units (IMUs) such as accelerometers that measure acceleration. Because of their small size and low cost, IMUs are now installed on most cell-phones. However, since velocity is estimated by integrating the measured acceleration, the measurement errors get accumulated over time, resulting in large drift errors. Although many methods have been proposed to address this limitation [2, 24], drift remains a fundamental problem of IMUs which limits their applicability in applications that require high accuracy and resolution.

The second class is visual sensors such as a camera that captures images of the surrounding world while the object (to which they are attached) moves. The sensor motion is estimated by measuring the motion of features in the images. Unfortunately, the image motion depends not only on the sensor motion but also the scene properties (e.g. depth or texture). In order to estimate the sensor motion, the effect of scene properties must be factored out. One popular approach to solve this problem is simultaneous localization and mapping (SLAM) [7], that simultaneously measures the 3D shape of the surrounding scene, as well as the 6 degrees-of-freedom (DOF) camera pose. Although SLAM based methods are implementable with only simple devices, they are computationally intensive and their performance is fundamentally limited by the surrounding scene's properties. For instance, while they can perform reliably if the scene has high frequency texture or sharp depth variations, their accuracy deteriorates if the scene has low frequency or repeated textures or small depth variations.

We propose a novel ego-motion sensor called *SpeDo* that addresses these fundamental limitations. SpeDo stands for Speckle defocus based odometer, and is a visual sensor based on a novel imaging method called *speckle defocus imaging*, where the surrounding scene is illuminated by a coherent light source (e.g., a laser), and imaged by a camera with a large defocus. Coherent light, on interacting with a scene, creates an interferometric intensity pattern on the sensor image plane, known as "speckle". The key theoretical insight underlying our work is that although the speckle pattern depends on scene characteristics (depths, reflectance properties), if the images are captured by a camera with large defocus, the *motion of speckle pattern is quasi-invariant to the scene characteristics*.

Theoretical model of speckle flow. We have developed a theoretical model of speckle flow (movement of speckle pattern in the captured images) due to camera and light source motion. We show that speckle flow is quasi-invariant to scene depths and textures, and depends only on the camera motion and focus plane's location. Moreover, the speckle pattern has a high spatial frequency, and even a small motion of the light source or the camera results in

large speckle movements. Thus, it is possible to achieve very high sensitivity and accuracy even with low cost off-the-shelf components. We also show that the speckle flow captured by a single camera can recover only four (out of 6) degrees of freedom of the motion information. But, it is possible to recover full 6 DOF by using speckle flow from two co-located cameras focused at different depths.

Hardware prototype and practical implications. Based on these theoretical results, we have built a hardware prototype of SpeDo that uses a laser source and two cameras focused at different depths. We have demonstrated that SpeDo can measure full 6 DOF absolute motion (not derivative of motion) for scenes with a broad range of shapes and textures. Since SpeDo is quasi-invariant to (and thus, does not need to estimate) scene properties, ego-motion can be measured simply by estimating optical flow at a few image locations (theoretically, only 2), which can be implemented in real time. We compare SpeDo with active visual SLAM (SLAM using active depth camera, e.g., a Kinect) and IMUs, and show that SpeDo achieves significantly higher resolution and accuracy. Because of the algorithmic simplicity, high accuracy, low cost of implementation and wide applicability, SpeDo can potentially become the method of choice in several applications requiring high speed ego-motion estimation in challenging real world scenarios where the scene properties are unknown.

Limitations. Because SpeDo uses active illumination, it has a shorter range than passive methods. Therefore, it can recover shorter trajectories (millimeter to meter scale) as compared to visual SLAM methods. An interesting future research direction would be to develop hybrid SpeDo and SLAM methods where SLAM is used for recovering large scale but coarse motion, and SpeDo is used to recover high resolution details at a small scale. The presence of bright ambient light (e.g., sunlight) can reduce the contrast of speckle pattern in captured images, thus lowering the SNR. This limitation is inherent in all active illumination methods, and can be addressed by using spectral filters or by concentrating light power [10]. Also, SpeDo cannot recover ego-motion if the scene is completely dark or consists of optically challenging materials e.g. translucent, transparent or irregular BRDF surfaces. This limitation is the same as that of optical laser mice. For such scenes, using combination of different sensors, such as SpeDo and IMUs can improve motion sensing quality. Finally, although speckle flow is quasi-invariant to scene depth, scenes with small scene depths that are comparable to the inter-camera distance in the hardware prototype may result in large estimation errors. This can be mitigated by building a setup where cameras and light source are exactly co-located by using beam-splitters.

2. Related Work

Ego-motion sensors. A principle similar to speckle defocus imaging is used in optical mice for measuring 2D ego-motion (planar translation). Recently, Zizka et al. developed a method for measuring 3D translation [30] by using speckle. Visual SLAM based approaches can recover full 6 DOF ego-motion, but have low accuracy and limited applicability since they require high-frequency scene texture or depths to perform reliably. IMUs are light-weight and low-cost, but suffer from the problem of large drift error. The proposed system can measure full 6 DOF absolute motion (not derivative of motion) with high accuracy, and is quasi-invariant to scene characteristics.

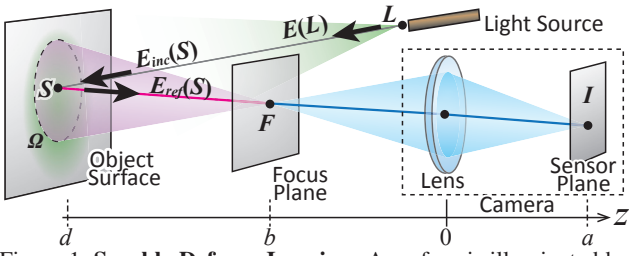


Figure 1. **Speckle Defocus Imaging.** A surface is illuminated by a coherent light source such as a laser. This creates speckle, a high frequency intensity distribution in 3D space due to interference of light. The surface is imaged by a camera with large defocus (camera's focus plane is distant from the surface). The intensity captured by the camera pixel $\boldsymbol{I}$ is the same as the speckle intensity at its conjugate point $\boldsymbol{F}$ on the focus plane.

Surface deformation measurement. Measuring surface deformation or strain fields is another popular application of speckle defocus imaging. If the sensor has a large defocus, even a small scene motion results in large speckle motion. This property has been used to develop high sensitivity deformation measurement methods [27, 4, 20, 14, 25]. Whereas these methods capture images with a single camera focus setting, a theoretical framework for analysing the effect of changing camera focus on speckle motion was developed by Gregory [9] and Hrabovsky et al. [13]. Based on this, Gregory [9] developed a method to measure 4 DOF (2D translation and 2D rotation) deformation by capturing images at multiple focus settings. Zhao et al. [29] extended it to measure 5 DOF (2D translation and 3D rotation) deformation. While these methods are used for measuring deformation for surfaces with known depths, our goal is different. We aim to develop a 6-DOF ego-motion sensor that can perform reliably even if the scene properties are unknown.

Other speckle based sensors. Speckle phenomena has been utilized in several other fields for measurement of various physical phenomena such as surface roughness[26, 18, 3], temperature[23], blood flow[6] and sound[28]. The roughness of a surface can change if it comes in contact with another surface. Based on this observation, [19] proposed a speckle based surface tampering detection method. Within computer vision, binocular stereo methods using laser speckle have also been proposed[15, 21, 16].

3. Speckle Defocus Imaging Model

Consider a surface illuminated by a coherent light source (e.g., laser), as shown in Fig. 1. Let the location of the point

light source be $\boldsymbol{L}$ and the wavelength of the light be λ. Let the electric field of the light emitted by the source[1] at a given time instant be given by the complex number $E(\boldsymbol{L})$, where $|E(\boldsymbol{L})|$ is the amplitude (square root of the source's intensity) and $arg(E(\boldsymbol{L}))$ is the initial phase at the light source.

Suppose the surface is imaged by a defocused camera (the focus plane of the camera is not on the surface). Let Ω be the surface patch imaged at a camera pixel location $\boldsymbol{I}$, as shown in Fig. 1. The size and shape of the patch is determined by the size of the camera defocus kernel. The electric field of light received at $\boldsymbol{I}$ is given by integrating the contributions from all the scene points $\boldsymbol{S}$ in the patch Ω:

$$E(\boldsymbol{I}) = \iint_{\Omega} E_{ref}(\boldsymbol{S}) \underbrace{e^{\left(\frac{2\pi i}{\lambda}\Gamma(\boldsymbol{S},\boldsymbol{I})\right)}}_{\text{Phase Transfer Function}} d\boldsymbol{S}, \quad (1)$$

where $E_{ref}(\boldsymbol{S})$ is the electric field of the light immediately after reflection from point $\boldsymbol{S}$ on the surface (at the same time instant)[2]. The phase transfer function from $\boldsymbol{S}$ to $\boldsymbol{I}$ differentiates speckle imaging (using coherent light) from conventional imaging (using incoherent light). The function is defined by $\Gamma(\boldsymbol{S},\boldsymbol{I})$, which is the optical path length between $\boldsymbol{S}$ and $\boldsymbol{I}$.

Let $\boldsymbol{S} \rightarrow \boldsymbol{I}$ be the light path between $\boldsymbol{S}$ and $\boldsymbol{I}$. All these paths (originating at points in region Ω) pass through point $\boldsymbol{F}$ on the camera's focus plane, where $\boldsymbol{F}$ is the conjugate point of pixel $\boldsymbol{I}$. Each of these paths can be divided into two sub-paths $\boldsymbol{S} \rightarrow \boldsymbol{F}$ and $\boldsymbol{F} \rightarrow \boldsymbol{I}$, as shown in red and blue colors, respectively, in Fig. 1.

Observation 1 *For all paths $\boldsymbol{S} \rightarrow \boldsymbol{I}$ originating at points $\boldsymbol{S} \in \Omega$, the optical length $\Gamma(\boldsymbol{F},\boldsymbol{I})$ of the $\boldsymbol{F} \rightarrow \boldsymbol{I}$ sub-path is constant.*

This is because the optical path length is the product of the geometric path length, and the refractive index of the medium. Since the rays that pass through $\boldsymbol{F}$ converge again at $\boldsymbol{I}$, their optical path lengths are the same [11]. Therefore, $\Gamma(\boldsymbol{S},\boldsymbol{I}) = \Gamma(\boldsymbol{S},\boldsymbol{F}) + \gamma$, where $\gamma = \Gamma(\boldsymbol{F},\boldsymbol{I})$ is a constant. Substituting in Eq. 1, we get:

$$E(\boldsymbol{I}) = \nu \iint_{\Omega} E_{ref}(\boldsymbol{S}) e^{\left(\frac{2\pi i}{\lambda}\Gamma(\boldsymbol{S},\boldsymbol{F})\right)} d\boldsymbol{S}, \quad (2)$$

where $\nu = e^{\left(\frac{2\pi i}{\lambda}\gamma\right)}$. Note that $|\nu| = 1$.

Focal Speckle. Next, we define *focal speckle* as the electric field distribution due to speckle on the camera focus plane. This is an important concept, and will be used repeatedly in the rest of the paper. Similar to Eq. 1, focal speckle $E(\boldsymbol{F})$ at a point $\boldsymbol{F}$ due to light paths between $\boldsymbol{S}$ and $\boldsymbol{I}$ is given as:

$$E(\boldsymbol{F}) = \iint_{\Omega} E_{ref}(\boldsymbol{S}) e^{\left(\frac{2\pi i}{\lambda}\Gamma(\boldsymbol{S},\boldsymbol{F})\right)} d\boldsymbol{S}. \quad (3)$$

Substituting Eq. 3 in Eq. 2, we get:

$$E(\boldsymbol{I}) = \nu E(\boldsymbol{F}). \quad (4)$$

[1]For ease of exposition, we assume an isotropic light source.

[2]$E_{ref}(\boldsymbol{S})$ can be calculated from the emitted light field $E(\boldsymbol{L})$, the surface reflectance term and the optical path length between $\boldsymbol{L}$ and $\boldsymbol{S}$. Please see the supplemental technical report for a complete definition.

The speckle image, i.e., the image brightness $U(\boldsymbol{I})$ measured at pixel $\boldsymbol{I}$ due to speckle is given as:

$$U(\boldsymbol{I}) = \xi |E(\boldsymbol{I})|^2 = \xi |E(\boldsymbol{F})|^2, \quad (5)$$

where ξ is a constant whose value depends on imaging parameters such as gain, aperture and exposure time. $|E(\boldsymbol{F})|^2$ is the intensity (square of amplitude) of the focal speckle. Hence, we get the following relationship between the speckle image captured by the camera and the focal speckle:

Result 1 (Speckle Image and Focal Speckle) *The brightness of the speckle image is equal (up to a constant scale) to the intensity of focal speckle.*

The above analysis and result are valid even if the camera focus plane is placed behind the image sensor (on the opposite side of the scene). In this case, the path length $\Gamma(\boldsymbol{F},\boldsymbol{I})$ is negative. We call this the *back focus configuration*. Such a configuration is rarely used in conventional imaging because the captured images have severe defocus blur. However, as we will show, the back focus configuration is important for speckle defocus imaging because the movement of speckle observed in the back focus images is different from that in front focus images.

Effect of ambient illumination. So far, we have assumed that the surface is illuminated only by a single coherent light source. In practice, the surface may be illuminated by additional uncontrollable non-coherent light sources, such as sunlight. Let $A(\boldsymbol{I})$ be the image brightness at pixel I due to illumination from such *ambient* light sources. The total image image brightness $T(\boldsymbol{I})$ is the sum of the ambient component and the speckle component: $T(\boldsymbol{I}) = U(\boldsymbol{I}) + A(\boldsymbol{I})$. The ambient component image $A(\boldsymbol{I})$ is given by the convolution of the camera defocus kernel and the surface texture. Since we consider cameras with a large defocus, $A(\boldsymbol{I})$ is severely blurred, and can be assumed to a constant Ψ such that $T(\boldsymbol{I}) = U(\boldsymbol{I}) + \Psi$. As a result, ambient illumination does not change the speckle image pattern, and for brevity, we do not consider it in the analysis for the rest of the paper.

4. Speckle Flow Model

In this section, we will derive a model for *speckle flow*, the local displacement of speckle pattern in the captured images due to camera and light source motion.

4.1. Speckle Flow Due To Camera Motion

Suppose a pixel $\boldsymbol{I}$ receives light rays reflected from a scene patch Ω, and that all these light rays pass through $\boldsymbol{F}$, the conjugate point of pixel $\boldsymbol{I}$, as shown in Fig. 2 (a). Next, suppose the camera moves (translates and rotates) while the light source and the scene surface remain fixed. If the camera motion is sufficiently small, point $\boldsymbol{F}$ remains on the focus plane of the camera [3], but becomes conjugate to a differ-

[3] Strictly speaking, $\boldsymbol{F}$ may not remain on the focus plane after camera motion. However, since the pixels have a finite size, the camera has a finite

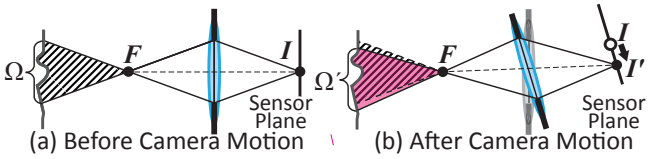


Figure 2. **Constancy of Speckle Intensity For Small Camera Motion.** (a) A pixel $\boldsymbol{I}$ receives light rays that are reflected from a surface patch Ω and pass through its conjugate point $\boldsymbol{F}$ on the focus plane. (b) If the camera moves by a small amount, point $\boldsymbol{F}$ remains on the focus plane, but becomes conjugate to a different pixel $\boldsymbol{I'}$, which receives light from scene patch Ω'. For small camera motion and large defocus, patches Ω and Ω' are approximately identical. As a result, intensity received at pixel $\boldsymbol{I'}$ is approximately the same as the intensity at $\boldsymbol{I}$ before camera motion.

ent pixel $\boldsymbol{I'}$, as shown in Fig. 2 (b). Pixel $\boldsymbol{I'}$ receives light from a different scene patch Ω'.

Observation 2 *For small camera motion and large defocus blur, patches Ω and Ω' are approximately identical, i.e., $\Omega \approx \Omega'$.*

Based on this observation, and following from Eq. 3, focal speckle $E'(\boldsymbol{F})$ after camera motion is given as:

$$E'(\boldsymbol{F}) = \iint_{\Omega'} E_{ref}(\boldsymbol{S})\, e^{\left(\frac{2\pi i}{\lambda}\Gamma(\boldsymbol{S},\boldsymbol{F})\right)} d\boldsymbol{S} \quad (6)$$

$$\approx \iint_{\Omega} E_{ref}(\boldsymbol{S})\, e^{\left(\frac{2\pi i}{\lambda}\Gamma(\boldsymbol{S},\boldsymbol{F})\right)} d\boldsymbol{S} \quad (7)$$

$$\approx E(\boldsymbol{F})\,. \quad (8)$$

where $E(\boldsymbol{F})$ is the focal speckle at point $\boldsymbol{F}$ before camera motion. Thus, we get the following result:

Result 2 (Constancy of Focal Speckle) *For small camera motion and large defocus blur, focal speckle (electric field distribution on the camera focus plane) remains approximately constant.*

Intuitively, we can think of point $\boldsymbol{F}$ as a fixed virtual scene point. For large defocus and small camera motion, the *brightness* of $\boldsymbol{F}$ (intensity of focal speckle at $\boldsymbol{F}$) remains approximately constant. This is true for all the points on the focus plane[4], even if the scene surface has strong depth discontinuity[5]. Note that the larger the Ω (large defocus blur), the longer the speckle pattern is preserved during camera motion.

Derivation of speckle flow. From Results **1** and **2**, it follows that the intensity received at pixel location $\boldsymbol{I'} = [u'\ v']$ after camera motion is approximately the same as the intensity at pixel location $\boldsymbol{I} = [u\ v]$ before camera motion. Thus, the speckle flow at $\boldsymbol{I}$ is given by the 2D image vector $[\Delta u\ \Delta v] = \boldsymbol{I'} - \boldsymbol{I}$. In the following, we derive the speckle flow at pixel $\boldsymbol{I}$ due to camera motion given by the translation and rotation vectors $\boldsymbol{t_C}$ and $\boldsymbol{\theta_C}$[6].

Suppose the origin of the camera coordinate system (CCS) is at the center of the lens, the X and Y axes are parallel to the image plane, and Z axis is along the optical axis. Let the coordinates of point $\boldsymbol{F}$ in the CCS before camera motion be given by the vector $\boldsymbol{F} = [x_F\ y_F\ z_F]^T$. Following Result **2**, since $\boldsymbol{F}$ can be treated as a fixed point in space, its coordinates in the CCS after camera motion are given by:

$$\boldsymbol{F'} = \boldsymbol{F} - \boldsymbol{t_C} + \mathbf{q}(-\boldsymbol{\theta_C})\,\boldsymbol{F}\,, \quad (9)$$

where $\mathbf{q}(\boldsymbol{\theta})$ is the 3×3 rotation matrix corresponding to the rotation vector $\boldsymbol{\theta}$ (the expression of $\mathbf{q}(\boldsymbol{\theta})$ is given in the supplementary technical report). Given coordinates $\boldsymbol{F}$ and $\boldsymbol{F'}$, and camera's projection matrix, image locations $\boldsymbol{I}$ and $\boldsymbol{I'}$ can be estimated by using perspective projection model (for details, see the supplementary technical report). Then, the camera speckle flow (speckle flow due to camera motion) $[\Delta u\ \Delta v] = \boldsymbol{I'} - \boldsymbol{I}$ is given as:

$$\begin{pmatrix} \Delta u \\ \Delta v \end{pmatrix} \approx \underbrace{\frac{-a}{pb}\begin{pmatrix} -1 & 0 & \hat{x_F} \\ 0 & -1 & \hat{y_F} \end{pmatrix}\boldsymbol{t_C}}_{\text{Focus Dependent}} - \underbrace{\frac{a}{p}\begin{pmatrix} 0 & -1 & +\hat{y_F} \\ 1 & 0 & -\hat{x_F} \end{pmatrix}\boldsymbol{\theta_C}}_{\text{Focus Invariant}} \quad (10)$$

where $\hat{x_F} = \frac{x_F}{z_F}$ and $\hat{y_F} = \frac{y_F}{z_F}$ are the normalized homogenous co-ordinates of point $\boldsymbol{F}$. p is camera's pixel size, a is the distance between lens and image sensor, and b is the distance between lens and focus plane, as shown in Fig. 1. Note that p, a and b are camera's intrinsic parameters, and hence, known a priori. The above equation is valid for each camera pixel location $[u, v]$. The terms Δu, Δv, $\hat{x_F}$ and $\hat{y_F}$ have (u, v) as arguments, i.e., $\Delta u(u, v)$, $\Delta v(u, v)$, $\hat{x_F}(u, v)$ and $\hat{y_F}(u, v)$. For the rest of the paper, we drop the argument (u, v) for brevity. The first term on the right hand side of Eq. 10 is the speckle flow cased by camera translation, and the second term is the speckle flow caused by camera rotation. The main characteristics of camera speckle flow are summarized in the following two results:

Result 3 (Depth Invariance Of Camera Speckle Flow) *Speckle flow due to small camera motion is invariant to d, the distance of the scene surface from the camera.*

Result 4 (Focus Dependence Of Camera Speckle Flow) *Speckle flow caused by camera rotation is invariant to the focus position. On the other hand, speckle flow caused by*

depth of field. In addition, the shape of focal speckle is like a 'cigar', with a finite length along the camera's optical axis [17]. Therefore, for small camera rotation, even if the DOF is small, the camera can observe the same speckle pattern at a point $\boldsymbol{F}$.

[4]This is reminiscent of the brightness constancy equation [12] used in optical flow, where brightness of scene points is assumed to remain constant for small camera motions. The important difference is that in optical flow, we consider real scene points, whereas in speckle defocus imaging, we consider virtual scene points.

[5] The scene points may have different intensity fall-offs, but since a pixel captures light from approximately the same set of points before and after motion, the speckle pattern remains constant.

[6]$\boldsymbol{\theta}$ is rotation vector whose direction is the rotation axis and magnitude is rotation angle.

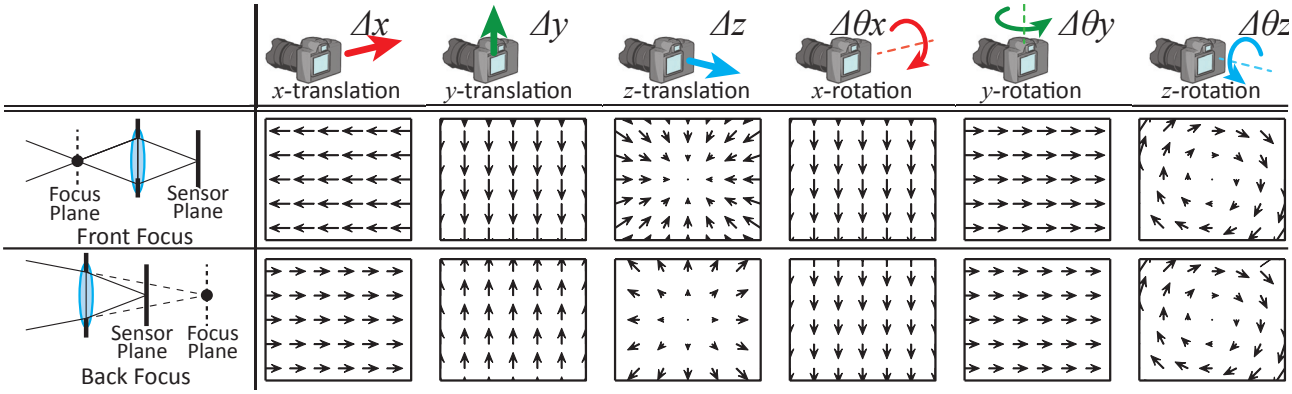


Figure 3. **Speckle Flow Patterns For Different Camera Motions.** We simulate the speckle flow field for different camera motions for both front and back focus settings. We assume that the lens has a long focal length. The flow fields can be divided into four categories - horizontal flow, vertical flow, zoom (in or out) and in-plane rotation. The flow due to camera translation (Δx, Δy, Δz) has opposite directions for front and back focus. In contrast, flow due to camera rotation ($\Delta\theta_x$, $\Delta\theta_y$, $\Delta\theta_z$) is in the same direction for front and back focus. This is an important property that will be used to distinguish camera rotation and translation.

camera translation depends on the focus position b. The direction of speckle flow due to translation is opposite for front and back focus configurations (focus plane in front of and behind the image sensor plane).

Fig. 3 shows the simulated speckle flow fields for different camera motions, both for front and back focus configurations (**please see videos on the project web-page [1]** for visualizations of speckle flow in videos of real scenes captured with a moving camera). Flow fields can be divided into 4 distinct categories - horizontal flow, vertical flow, zoom (in or out), and in-plane rotation. The speckle flow due to z-translation and z-rotation is smaller than that of the other four motions. Hence, we magnified these two speckle fields for visualization. Notice that flow due to camera translation (Δx, Δy, Δz) has opposite directions for front and back focus. In contrast, flow due to camera rotation ($\Delta\theta_x$, $\Delta\theta_y$, $\Delta\theta_z$) is in the same direction for front and back focus. This is an important property that we will use to distinguish speckle flow fields due to camera rotation and translation.

Next, we will derive the source speckle flow (speckle flow due to light source motion) while the scene surface and camera remain stationary. Suppose the point source moves from location $\boldsymbol{L}$ to $\boldsymbol{L'}$. Suppose E and E' are the focal speckle fields before and after the source motion, respectively. We use a result from the optics literature that shows that if the light source motion is small, the focal speckle field before camera motion at a point $\boldsymbol{F}$ *is the same* as the focal speckle field after camera motion at another point $\boldsymbol{F'}$ on the focus plane, i.e., $E'(\boldsymbol{F'}) = E(\boldsymbol{F})$ [5]. This is illustrated in Figure 4. In the following, we derive the relationship between $\boldsymbol{F}$ and $\boldsymbol{F'}$, and use that to derive the expression for source speckle flow.

Let $\boldsymbol{S}$ be the surface point on the line joining camera center and original focus point $\boldsymbol{F}$. Let $\boldsymbol{sl}$ and $\boldsymbol{sl'}$ be the unit vectors in the directions $\overrightarrow{\boldsymbol{SL}}$ and $\overrightarrow{\boldsymbol{SL'}}$, respectively. Let $\boldsymbol{\Delta sl} = \boldsymbol{sl'} - \boldsymbol{sl}$ be the change in unit vector direction from $\boldsymbol{S}$ to light source. Similarly, let $\boldsymbol{sf}$ and $\boldsymbol{sf'}$ be the unit vectors

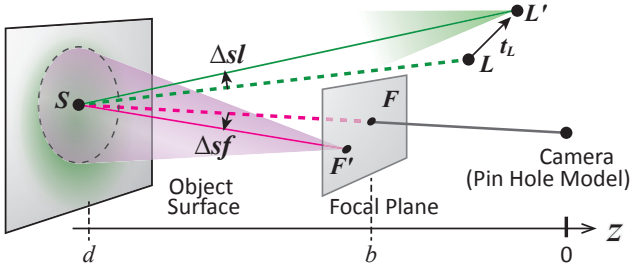


Figure 4. **Movement Of Speckle Field Due To Light Source Motion.** If the light source moves by a small amount from $\boldsymbol{L}$ to $\boldsymbol{L'}$, the speckle intensity at a focal point $\boldsymbol{F}$ before motion is the same as the intensity at a different focal point $\boldsymbol{F'}$ after motion.

in the directions $\overrightarrow{\boldsymbol{SF}}$ and $\overrightarrow{\boldsymbol{SF'}}$, and $\boldsymbol{\Delta sf} = \boldsymbol{sf'} - \boldsymbol{sf}$ be the change in the unit vector direction from $\boldsymbol{S}$ to the focal point. Then, $\boldsymbol{\Delta sf}$ and $\boldsymbol{\Delta sl}$ are related as [5]:

$$\boldsymbol{\Delta sf} + \boldsymbol{\Delta sl} = 0 \tag{11}$$

By using the above equation, we can determine $\boldsymbol{F'}$ by intersecting the ray along the unit vector $\boldsymbol{sf'}$ with the focus plane. Then, by using the perspective projection model and paraxial approximation, we can determine the image pixel locations $\boldsymbol{I}$ and $\boldsymbol{I'}$ corresponding to focal points $\boldsymbol{F}$ and $\boldsymbol{F'}$, respectively (for derivation, see the supplementary technical report). Then, the source speckle flow $[\Delta u\ \Delta v] = \boldsymbol{I'} - \boldsymbol{I}$ is given as:

$$\begin{pmatrix} \Delta u \\ \Delta v \end{pmatrix} = \frac{-a}{p}\left(\frac{1}{b} - \frac{1}{d}\right)\begin{pmatrix} -1 & 0 & \hat{x_F} - \frac{x_L}{d} \\ 0 & -1 & \hat{y_F} - \frac{y_L}{d} \end{pmatrix} \boldsymbol{t_L} \tag{12}$$

4.2. Speckle Flow Due To Source Motion

where x_L, y_L are the x, y coordinates of the light source (before movement) in the CCS. The rest of the terms are as defined in the previous sub-section (before and after Eq. 10). In this paper, we assume that the focus plane is significantly closer to the camera than the scene surface, i.e., $\frac{1}{|b|} \gg \frac{1}{|d|}$ and that the camera-source distance is negligible,

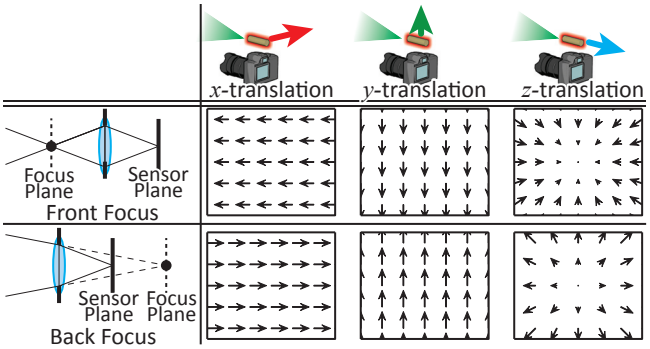


Figure 5. **Speckle Flow Patterns For Different Light Source Motions.** Speckle flow fields for different light source motions for both front and back focus settings. Flow fields for back and front focus have different directions.

i.e., $\frac{x_L}{d} \approx 0$ and $\frac{y_L}{d} \approx 0$. Under these approximations, the source speckle flow is given as:

$$\begin{pmatrix} \Delta u \\ \Delta v \end{pmatrix} \approx \frac{-a}{bp} \begin{pmatrix} -1 & 0 & \hat{x_F} \\ 0 & -1 & \hat{y_F} \end{pmatrix} \boldsymbol{t_L} \qquad (13)$$

The main characteristics of source speckle flow are summarized in the following two results:

Result 5 (Quasi Depth Invariance Of Source Speckle Flow) *Under assumptions of large camera defocus, large scene depth and small camera-source distance, speckle flow due to small source motion is quasi-invariant to scene depth d.*

Result 6 (Focus Dependence Of Source Speckle Flow) *The direction of source speckle flow is opposite for front and back focus configurations.*

Fig. 5 shows the simulated speckle flow fields for different source motions (see videos on the project web-page [1] for visualizations of source speckle flow in videos). In accordance with Result **6**, speckle flow fields have opposite directions for front and back focus.

5. Ego-Motion Recovery From Speckle Flow

In this section, we present our method for ego-motion recovery from speckle flow. We assume that the light source and the camera are fixed with respect to each other, and move together as a single unit, called a SpeDo. We assume that a SpeDo's coordinate system is the same as its camera's coordinate system.

Let the translation and rotation of a SpeDo be given by the vectors $\boldsymbol{t_S}$ and $\boldsymbol{\theta_S}$. Since the SpeDo's coordinate system is the same as the CCS, the translation and rotation of the camera are the same as that of the SpeDo, i.e., $\boldsymbol{t_C} = \boldsymbol{t_S}$ and $\boldsymbol{\theta_C} = \boldsymbol{\theta_S}$. The translation of the light source is given as:

$$\boldsymbol{t_L} = \boldsymbol{t_S} + \mathbf{q}(\boldsymbol{\theta_S})\boldsymbol{L}\,, \qquad (14)$$

where $\boldsymbol{L}$ is the location of the source in the CCS, and $\mathbf{q}(\boldsymbol{\theta_S})$ is the rotation matrix corresponding to the rotation vector $\boldsymbol{\theta_S}$. Since we assume that the source-camera distance is negligible, i.e., $\boldsymbol{L} \approx 0$, we approximate $\boldsymbol{t_L} \approx \boldsymbol{t_S}$.

Under the assumption of small motion, the total speckle flow is the sum of the camera speckle flow (Eq. 10) and the source speckle flow (Eq. 13):

$$\begin{pmatrix} \Delta u \\ \Delta v \end{pmatrix} \approx \frac{-2a}{pb} \begin{pmatrix} -1 & 0 & \hat{x_F} \\ 0 & -1 & \hat{y_F} \end{pmatrix} \boldsymbol{t_S} - \frac{a}{p} \begin{pmatrix} 0 & -1 & +\hat{y_F} \\ 1 & 0 & -\hat{x_F} \end{pmatrix} \boldsymbol{\theta_S} \qquad (15)$$

The above equation represents the SpeDo speckle flow (speckle flow due to the motion of a SpeDo). We can write the above as a linear system of equations:

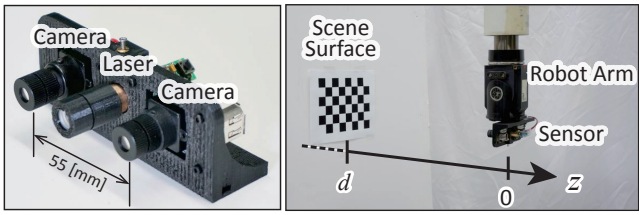
$$\begin{pmatrix} \Delta u \\ \Delta v \end{pmatrix} = \mathbf{M} \begin{pmatrix} \boldsymbol{t_S} \\ \boldsymbol{\theta_S} \end{pmatrix}, \qquad (16)$$

where $\mathbf{M}$ is the (known) 2×6 measurement matrix. This system has 2 equations and 6 unknowns, and hence, underconstrained. However, since the above equation applies individually to every pixel location in the captured image, we can increase the number of equations by considering speckle flow at multiple pixel locations. Specifically, if flow at P different pixels is used, we can obtain $2P$ equations.

Degrees of freedom in a single speckle flow. From the above discussion, it may appear that by combining speckle flow information from $P \geq 3$ pixels, we can recover the full 6 DOF motion. However, the equations from different pixels are not all independent. Specifically, a single speckle flow field over an image has only 4 degrees of freedom. Intuitively, this is because there are only four different kinds of speckle flow (as discussed in the previous section) - horizontal translation, vertical translation, zoom, and rotation. Therefore, speckle flow computed for a single camera can recover only 4 degrees of freedom. How can we recovery the full 6 DOF motion information?

Bi-Focal SpeDo. The key idea is that since speckle flow for front and back focus configurations are different, by combining information from two speckle flows, one with front focus and the other with back focus, we can recover the full 6 DOF motion information. For example, x-translation and y-rotation of the camera both produce horizontal speckle flow (Fig. 3), and thus, cannot be distinguished from a single flow field. But, since the speckle flow directions are different for front and back focus configurations, x-translation and y-rotation can be distinguished by using both front and back focus speckle flows. Based on this, we propose Bi-Focal SpeDo, a system that uses two co-located cameras with front and back focus configurations.

Ego-motion recovery algorithm. Our ego-motion recovery algorithm consists of computing speckle flow (by using optical flow) for the two cameras of a Bi-Focal SpeDo system. The speckle flow values from both cameras are collected into a single linear system, as given in Eq. 16. The size of matrix $\mathbf{M}$ is $4P \times 6$, where P is the number of pixel



(a) Hardware Prototype (b) Experiment Setup

Figure 6. **Experimental Setup.** (a) Hardware prototype of the proposed SpeDo system consisting of two cameras, one with a front focus setting and the other with a back focus setting, and a laser source. (b) In order to measure the accuracy of SpeDo, we mounted the prototype on a robot arm and applied a variety of known motions to it. We used a wide range of scenes, including a flat white plane, a textured plane, and a scene consisting of a variety of objects of different shapes and textures.

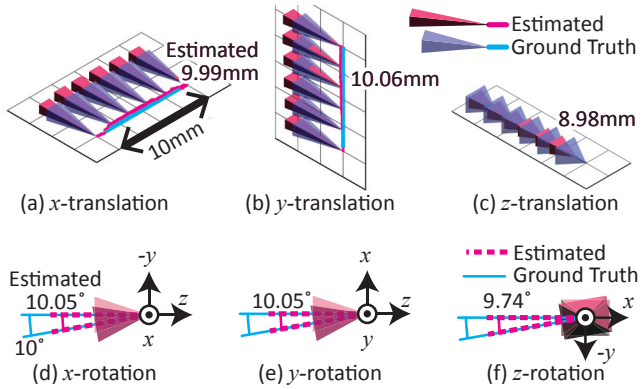


Figure 7. **6 DOF Ego-Motion Measurement Using SpeDo.** Ground truth and measured trajectories for six different motions (translations and rotations along three axes). The range of the translation and rotation trajectories is 10 mm and 10°, respectively. SpeDo recovers every trajectory with high accuracy. The sensitivity of estimation for translation and rotation along z axis is lower than the other two axes, resulting in lower accuracy.

locations whose flow is used (each pixel provides 4 equations, two in each speckle flow). This system is solved using linear least squares: $\begin{pmatrix} \boldsymbol{t_S} \\ \boldsymbol{\theta_S} \end{pmatrix} = \mathbf{M}^{\dagger} \begin{pmatrix} \Delta u \\ \Delta v \end{pmatrix}$ where $\dagger$ is the pseudo-inverse. Theoretically, speckle flow values from both cameras for $P = 2$ pixel locations are sufficient to recover the motion parameters. In practice, for higher robustness, we use a larger number of pixels (e.g., 25).

6. Hardware Prototype And Results

Our hardware prototype consists of two Point Grey Research FireFly MV cameras with 25mm F2.0 lenses, and a green 532 nm laser pointer, as shown in Fig. 6 (a). Ideally, the cameras and the laser should be placed at exactly the same position, which can be achieved by using beam-splitters. In our implementation, the cameras and the source are placed adjacent to each other with a small separation, which may result in estimation errors. In order to address this problem, we use the depth dependent version of source

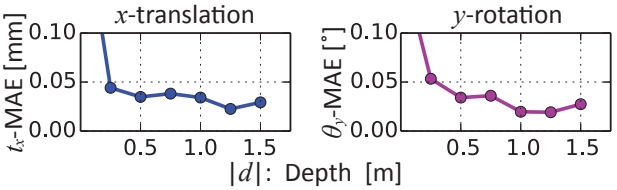


Figure 8. **Estimation Error Vs. Scene Depths.** In order to demonstrate the effect of scene depths, we performed ego-motion estimation with the scene (a single fronto-parallel plane) placed at different scene depths, and computed the mean error for each depth. Error plots for two different trajectories (1mm translation and 1° rotation) are shown in (a) and (b). When scene depth is larger than 0.5 meters, the mean absolute error is less than 0.05 mm and 0.05°. However, errors are larger if the depth is smaller than 0.5 meters.

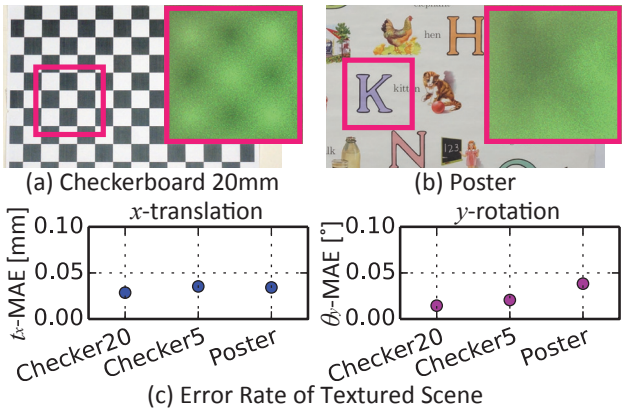
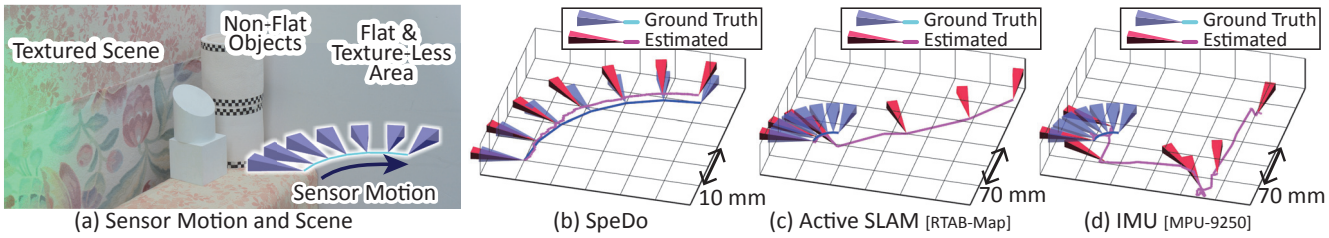


Figure 9. **Estimation Error Vs. Scene Textures.** (a-b) We performed ego-motion estimation with the scene (a fronto-parallel plane) having various textures including two checker board patterns with checkers of different sizes and a poster with several images. Insets show captured images. Due to large defocus, the texture is almost completely blurred, making SpeDo quasi-invariant to scene texture. (c) Plots of mean error for two different trajectories (translation and rotation). In contrast, SpeDo achieves a low error rate irrespective of the surface texture.

speckle flow (Eq. 12 instead of Eq. 13) and use an approximate scene depth $d = 1.0$ meter (assumed to be constant for the entire scene). This is not a fundamental requirement for the method. If the cameras and the source are co-located, we simply use the depth invariant version of speckle flow equation. The blur size, or focus position b, should be chosen to satisfy following conditions. First, the focus settings are chosen so that $\frac{1}{|b|} \gg \frac{1}{|d|}$ in order to achieve quasi depth invariance (Eq. 12). Second, the blur size should be sufficiently large to ensure focal speckle constancy (Result 2) during camera motion [7]. Third, the blur kernel should be at least twice as large as the speckle size so that the high frequency background texture is blurred and only speckle pattern is observed. In our implementation, we choose

[7] The relationship between blur size and the duration for which speckle remains constant during camera motion (known as correlation length) is well analyzed in the optics community [8].



(a) Sensor Motion and Scene (b) SpeDo (c) Active SLAM [RTAB-Map] (d) IMU [MPU-9250]

Figure 10. **Comparisons with Visual SLAM and IMUs.** Results of comparisons between active visual SLAM (using a v2 Kinect), IMUs and SpeDo for a trajectory containing both rotation and translation. The scene consists of a variety of objects with different scene depths and textures. An IMU measures acceleration which must be integrated twice for estimating the sensor position. Consequently, small errors in the measured acceleration result in large position errors, even if the trajectory is relatively small. The positions measurements from using SLAM have large errors, especially in the second half of the trajectory where the camera images the textureless and planar portion of the scene. In contrast, SpeDo measures the camera pose with high accuracy over the entire trajectory.

$b = \pm 0.2$ meter, because of hardware limitation. Ideally, shorter b (larger defocus) will achieve better results. In order to compute the optical flow between speckle images, we use the phase only correlation algorithm [22].

6 DOF ego-motion measurement using SpeDo. To demonstrate the ability of SpeDo to measure 6 DOF ego-motion, we mounted our hardware prototype on a robot arm and applied a variety of known motion trajectories to it. The scene was a textureless flat plane 1.0 meter away from the sensor, as shown in Fig.6 (b). Fig. 7 shows the ground truth trajectories and measured trajectories for six different motions (translations and rotations along three axes). In all cases, SpeDo recovers the trajectory with high accuracy. We calculate the mean absolute error (MAE) as $\text{mean}\,|t_{gt} - t_m|$ and $\text{mean}\,|\theta_{gt} - \theta_m|$ for the translation and rotation motions, where t_{gt} (θ_{gt}) and t_m (θ_{gt}) are the ground truth and measured translation (rotation), respectively. The MAE of translations (per 1 mm translation) are x : 0.026 mm, y : 0.026 mm and z : 0.16, and the MAE of rotation (per 1° rotation) are θ_x : 0.023°, θ_y : 0.020° and θ_z : 0.082°. The sensitivity of estimation for translation and rotation along z axis is lower than the other two axes, resulting in lower accuracy.

Experiments to demonstrate quasi-invariance to scene depth. Theoretically, SpeDo is quasi-invariant to scene depths. However, in our hardware prototype, since the camera and the source are not exactly co-located (distance between cameras is 55 mm), the measurement accuracy is low for scene depths that are comparable to the inter-camera distance. In order to demonstrate the effect of scene depths, we performed ego-motion estimation with the scene (a single fronto-parallel plane) placed at different scene depths between 1.5 meters and 0.125 meters, and measured the MAE for each scene depth. Example error plots for two different trajectories (1 mm translation and 1° rotation) are shown in Fig. 8 (for more results, **please see the project web-page [1]**). When scene depth is larger than 0.5 meters, the error is less than 0.05 mm and 0.05°. However, errors are larger if the depth is smaller than 0.5 meters. This limitation can be addressed by using a setup where the light source and cameras are co-located by using beam-splitters.

Experiments to demonstrate invariance to scene texture. We performed ego-motion estimation with the scene (a single plane at depth of 0.75 meter) having various textures with a wide range of spatial frequencies. Fig. 9 shows the results for three textures - two checker board patterns with checkers of size 20 mm and 5 mm, and a poster with several images. Note that the checker boards patterns have repeated textures, which cause passive methods (such as SLAM) that rely on feature matching to produce erroneous results. In contrast, SpeDo measures ego-motion with a low error rate.

Comparisons with Active SLAM and IMUs for complex scene. We compared the performance of SpeDo with active visual SLAM (SLAM using an active depth camera, e.g., a Kinect) and IMUs for a variety of motion trajectories containing both translation and rotation. For comparisons with active SLAM, we used the RTAB-Map with Kinect v2. The IMU used for comparisons is an Invensens MPU-9250 sensor, which contains an accelerometer, a gyroscope and a magnetometer. Fig 10 shows the comparison results for an example trajectory. The scene contains a variety of objects with different scene depths and textures. Active SLAM results in large errors in the second half of the trajectory where the camera images the textureless and planar portion of the scene. An IMU measures acceleration which must be integrated twice for estimating the sensor position. Consequently, small errors in the measured acceleration result in large position errors, even if the trajectory is relatively small. In contrast, SpeDo measures the camera pose with high accuracy over the entire trajectory.

Measuring complex 6 DOF motions. We have used SpeDo to measure a variety of complex motion trajectories, including those with sharp gradients and self-intersections, for example, roman numerals and hand motions. The results are shown in videos on the project web-page [1]. In all cases, SpeDo recovers the ego-motion with high accuracy.

Acknowledgment. This research was conducted in the Computer Vision Laboratory at Columbia University. Kensei Jo is with Sony Corporation and was a visiting researcher to Columbia University. Shree Nayar received a letter of appreciation with a prize from Sony Corporation for his contributions to research and education.

References

[1] CAVE | Computer Vision Laboratory - Columbia University. http://www.cs.columbia.edu/CAVE/projects/spedo/. 5, 6, 8

[2] T. Beravs, J. Podobnik, and M. Munih. Three-Axial Accelerometer Calibration Using Kalman Filter Covariance Matrix for Online Estimation of Optimal Sensor Orientation. *IEEE Transactions on Instrumentation and Measurement*, 61(9):2501–2511, Sept. 2012. 1

[3] J. D. R. Buchanan, R. P. Cowburn, A.-V. Jausovec, D. Petit, P. Seem, G. Xiong, D. Atkinson, K. Fenton, D. a. Allwood, and M. T. Bryan. Forgery: 'fingerprinting' documents and packaging. *Nature*, 436(7050):475, July 2005. 2

[4] D. J. Chen, F. P. Chiang, Y. S. Tan, and H. S. Don. Digital speckle-displacement measurement using a complex spectrum method. *Applied optics*, 32(11):1839–49, Apr. 1993. 2

[5] F. Chiang and D. Li. Laws of laser speckle movement in space. *Optical Engineering*, 25:667–670, 1986. 5

[6] A. K. Dunn, H. Bolay, M. A. Moskowitz, and D. A. Boas. Dynamic imaging of cerebral blood flow using laser speckle. *Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism*, 21(3):195–201, 2001. 2

[7] M. W. M. Gamini Dissanayake, P. Newman, S. Clark, H. F. Durrant-Whyte, and M. Csorba. A solution to the simultaneous localization and map building (SLAM) problem. *IEEE Transactions on Robotics and Automation*, 17(3):229–241, 2001. 1

[8] B. Gombkötő and J. Kornis. Success rate and speckle correlation in electronic speckle photography. *Optics communications*, 201(4):289–292, 2002. 7

[9] D. A. Gregory. Basic physical principles of defocused speckle photography: a tilt topology inspection technique. *Optics & Laser Technology*, (October):201–213, 1976. 2

[10] M. Gupta, Q. Yin, and S. K. Nayar. Structured Light in Sunlight. *2013 IEEE ICCV*, 2013. 2

[11] E. Hecht and A. Zajac. *Optics*. 3

[12] B. K. P. Horn and B. G. Schunck. Determining optical flow. *ARTIFICAL INTELLIGENCE*, 17:185–203, 1981. 4

[13] M. Hrabovsky, Z. Bac, and P. Horva. Theory of speckle displacement and decorrelation and its application in mechanics. 32:395–403, 2000. 2

[14] C. Joenathan and H. J. Tiziani. Speckle and Speckle Metrology, 2007. 2

[15] L. Larsson, M. Sjödahl, and F. Thuvander. Microscopic 3-D displacement field measurements using digital speckle photography. *Optics and Lasers in Engineering*, 41(5):767–777, May 2004. 2

[16] P. F. Luo, Y. J. Chao, M. a. Sutton, and W. H. Peters. Accurate measurement of three-dimensional deformations in deformable and rigid bodies using computer vision. *Experimental Mechanics*, 33(2):123–132, June 1993. 2

[17] H. J. Rabal and R. A. Braga. *Dynamic Laser Speckle and Applications*. 2010. 4

[18] A. Sharma, L. Subramanian, and E. A. Brewer. PaperSpeckle: microscopic fingerprinting of paper. In *Proceedings of the 18th ACM conference on Computer and communications security - CCS '11*, page 99, 2011. 2

[19] Y. C. Shih, A. Davis, S. W. Hasinoff, F. Durand, and W. T. Freeman. Laser speckle photography for surface tampering detection. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 33–40, 2012. 2

[20] M. Sjödahl and H. O. Saldner. Three-dimensional deformation field measurements with simultaneous TV holography and electronic speckle photography. *Applied optics*, 36(November 1996):3645–3648, 1997. 2

[21] P. Synnergren and M. Sjödahl. A stereoscopic digital speckle photography system for 3-D displacement field measurements. *Optics and Lasers in Engineering*, 31, 1999. 2

[22] K. Takita, T. Aoki, Y. Sasaki, T. Higuchi, and K. Kobayashi. High-Accuracy Subpixel Image Registration Based on Phase-Only Correlation. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E86-A(8):1925–1934, 2003. 8

[23] V. Trivedi, S. Mahajan, V. Chhaniwal, Z. Zalevsky, B. Javidi, and A. Anand. Optical temperature sensor using speckle field. *Sensors and Actuators A: Physical*, 216:312–317, Sept. 2014. 2

[24] S.-h. P. Won and F. Golnaraghi. A Triaxial Accelerometer Calibration Method Using a Mathematical Model, 2010. 1

[25] I. Yamaguchi. Holography, speckle, and computers. *Optics and Lasers in Engineering*, 39(4):411–429, Apr. 2003. 2

[26] I. Yamaguchi, K. Kobayashi, and L. Yaroslavsky. Measurement of surface roughness by speckle correlation, Nov. 2004. 2

[27] I. Yamaguchi and H. Saito. Deformation measurement by speckle photography. *High Speed Photography and Photonics*, 1979. 2

[28] Z. Zalevsky, Y. Beiderman, I. Margalit, S. Gingold, M. Teicher, V. Mico, J. Garcia, V. M. Zeev Zalevsky, Yevgeny Beiderman, Israel Margalit, Shimshon Gingold, Mina Teicher, and J. Garcia. Simultaneous remote extraction of multiple speech sources and heart beats from secondary speckles pattern. *Optics express*, 17(24):21566–21580, 2009. 2

[29] W. Zhao and J. Wu. A novel optical technique for measuring 5-axis surface movement. *Proceedings of the SPIE*, 5606:66–73, Dec. 2004. 2

[30] J. Zizka, A. Olwal, and R. Raskar. SpeckleSense: fast, precise, low-cost and compact motion sensing using laser speckle. *Proceedings ACM Symposium on User Interface Software and Technology (UIST)*, page 10, 2011. 2

Supplementary Technical Report for SpeDo

Kensei Jo Mohit Gupta Shree K. Nayar

Columbia University

Abstract

In this report, we provide the derivations of the speckle flow model and additional results of ego-motion estimation using SpeDo.

1 Camera Coordinate System

We define the camera coordinate system (CCS) as shown in Fig. 1. The origin is at the lens center and the direction of z axis is towards the image sensor plane. Let the 3D location of a scene point $\boldsymbol{F}$ be given by the vector (x_F, y_F, z_F). Suppose $\boldsymbol{F}$ is imaged at a pixel location (u_F, v_F) on the sensor plane. Then, (x_F, y_F, z_F) and (u_F, v_F) are related as:

$$s \begin{pmatrix} u_F \\ v_F \\ 1 \end{pmatrix} = \mathbf{A} \begin{pmatrix} x_F \\ y_F \\ z_F \end{pmatrix}. \tag{1}$$

where s is an arbitrary scale factor, and $\mathbf{A}$ is the camera's 3×3 intrinsic matrix:

$$\mathbf{A} = \begin{pmatrix} \frac{-a}{p} & 0 & c_u \\ 0 & \frac{-a}{p} & c_v \\ 0 & 0 & 1 \end{pmatrix}, \ \mathbf{A}^{-1} = \begin{pmatrix} \frac{p}{-a} & 0 & \frac{p}{a} c_u \\ 0 & \frac{p}{-a} & \frac{p}{a} c_v \\ 0 & 0 & 1 \end{pmatrix}, \tag{2}$$

a is the distance between the image sensor and the lens, (c_u, c_v) are the coordinates of the principal point on the image plane, and p is the pixel size of image sensor.

The homogeneous coordinates $(\hat{x_F}, \hat{y_F}, 1) = (\frac{x_F}{z_F}, \frac{x_F}{z_F}, 1)$ are related to the pixel location (u_F, v_F) as:

$$\begin{pmatrix} \hat{x_F} \\ \hat{y_F} \\ 1 \end{pmatrix} = \mathbf{A}^{-1} \begin{pmatrix} u_F \\ v_F \\ 1 \end{pmatrix}. \tag{3}$$

2 Rotation Matrices For Small Rotations

In this section, we will derive the expression for the rotation matrix that is used to express the change in coordinates of 3D points due to rotation of the coordinate system. Consider a point with 3D coordinates $\boldsymbol{P} = (P_x,\ P_y,\ P_z)^T$ in space. Suppose the coordinate system rotates so that the rotation is given by the vector $\boldsymbol{\theta} = (\theta_x,\ \theta_y,\ \theta_z)^T$, where the amount (angle) of rotation is equal to the magnitude $|\boldsymbol{\theta}|$, and the axis of rotation

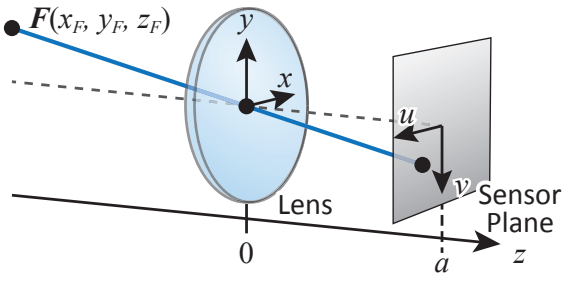


Figure 1: **Camera Coordinate System.**

is the unit vector $\frac{\boldsymbol{\theta}}{|\boldsymbol{\theta}|}$. If the amount of rotation $|\boldsymbol{\theta}|$ is small, the coordinates of $\boldsymbol{P}$ after the rotation are given by:

$$\boldsymbol{P_{rot}} = \boldsymbol{P} + \mathbf{q}(\boldsymbol{\theta})\boldsymbol{P}, \tag{4}$$

where the rotation matrix $\mathbf{q}(\boldsymbol{\theta})$ is given by:

$$\mathbf{q}(\boldsymbol{\theta}) \ = \ \begin{pmatrix} 0 & -\theta_z & +\theta_y \\ +\theta_z & 0 & -\theta_x \\ -\theta_y & +\theta_x & 0 \end{pmatrix}. \tag{5}$$

Note that the matrix $\mathbf{q}(\boldsymbol{\theta})$ has the same form as a cross-product matrix. Hence, the product $\mathbf{q}(\boldsymbol{\theta})\boldsymbol{P}$ can also be written as a cross-product: $\mathbf{q}(\boldsymbol{\theta})\boldsymbol{P} = \boldsymbol{\theta} \times \boldsymbol{P}$. We can replace the order of the operators in a cross-product, and write this expression as:

$$\mathbf{q}(\boldsymbol{\theta})\boldsymbol{P} = \boldsymbol{\theta} \times \boldsymbol{P} = -\boldsymbol{P} \times \boldsymbol{\theta} = \mathbf{q}(-\boldsymbol{P})\boldsymbol{\theta}. \tag{6}$$

This is an important property, and will be used in further derivations.

3 Coherent Light After Reflection

In this section, we will derive the $E_{ref}(\boldsymbol{S})$, which is the electric field of the coherent light immediately after reflection from point $\boldsymbol{S}$ on the surface.

Consider a surface illuminated by a coherent light source (e.g., laser), as shown in Fig. 2. Let the location of the point light source be $\boldsymbol{L}$ and the wavelength of the emitted light be λ. Let the electric field of the light emitted by the source[1] at a given time instant be given by the complex number $E(\boldsymbol{L})$, where $|E(\boldsymbol{L})|$ is the amplitude (square root of the source's intensity) and $arg(E(\boldsymbol{L}))$ is the initial phase at the light source.

[1] For ease of exposition, we assume an isotropic light source.

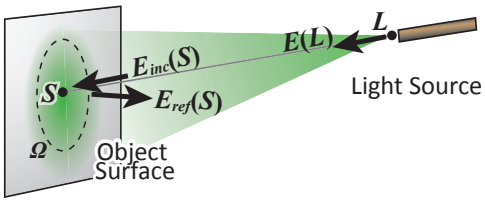


Figure 2: **The Electric Field After Reflection:** The point light source at location $\boldsymbol{L}$ emits coherent light and the object surface refracts the light. The reflected light at location $\boldsymbol{S}$ can be calculated from the distance between $\boldsymbol{L}$ and $\boldsymbol{S}$, and the reflectance value $\alpha(\boldsymbol{S})$.

As light travels from the source to the surface, the electric field's amplitude decreases (due to inverse square intensity fall-off from a point source), and the phase varies according to the travel distance. The electric field of the light incident at a point $\boldsymbol{S}$ on the surface (at the same time instant) is related to the emitted light field $E(\boldsymbol{L})$ as:

$$E_{inc}(\boldsymbol{S}) = \frac{E(\boldsymbol{L})}{\Gamma(\boldsymbol{L},\boldsymbol{S})} e^{\left(\frac{2\pi i}{\lambda}\Gamma(\boldsymbol{L},\boldsymbol{S})\right)}, \tag{7}$$

where $\Gamma(\boldsymbol{L},\boldsymbol{S})$ is the optical path length between $\boldsymbol{L}$ and $\boldsymbol{S}$. The electric field of the light immediately after reflection from point $\boldsymbol{S}$ is given by:

$$E_{ref}(\boldsymbol{S}) = \alpha(\boldsymbol{S}) E_{inc}(\boldsymbol{S}) = \alpha(\boldsymbol{S}) \frac{E(\boldsymbol{L})}{\Gamma(\boldsymbol{L},\boldsymbol{S})} e^{\left(\frac{2\pi i}{\lambda}\Gamma(\boldsymbol{L},\boldsymbol{S})\right)}, \tag{8}$$

where $\alpha(\boldsymbol{S})$ is the surface reflectance term at point $\boldsymbol{S}$. It encapsulates the foreshortening effect and the BRDF of $\boldsymbol{S}$.

4 Speckle Flow Due To Camera Motion

In this section, we will derive the speckle flow (motion of speckle in captured images) due to camera motion given by the translation and rotation vectors $\boldsymbol{t_C}$ and $\boldsymbol{\theta_C}$. Let the coordinates (in camera coordinate system) of a point $\boldsymbol{F}$ on camera focus plane before camera motion be given by the vector $\boldsymbol{F} = (x_F,\ y_F,\ z_F)^T$. Following Result **2** in the paper, since $\boldsymbol{F}$ can be treated as a fixed point in space, its coordinates in the CCS after camera motion are given by:

$$\boldsymbol{F'} = \boldsymbol{F} - \boldsymbol{t_C} + \mathbf{q}(-\boldsymbol{\theta_C})\,\boldsymbol{F}, \tag{9}$$

By using Eq. 6, we change the order of operators on the right hand side of the above equation, and get:

$$\boldsymbol{F'} = \boldsymbol{F} - \boldsymbol{t_C} + \mathbf{q}(\boldsymbol{F})\,\boldsymbol{\theta_C}. \tag{10}$$

Next, let the vector $\Delta\boldsymbol{F} = \boldsymbol{F'} - \boldsymbol{F} = (\Delta x_F,\ \Delta y_F,\ \Delta z_F)^T$ denote the motion of point $\boldsymbol{F}$ in CCS due to camera motion:

$$\Delta\boldsymbol{F} = -\boldsymbol{t_C} + \mathbf{q}(\boldsymbol{F})\,\boldsymbol{\theta_C}. \tag{11}$$

Computing image locations before and after motion. Let $\boldsymbol{F}$ be imaged at a pixel location (u,v) on the sensor plane before camera motion. By simplifying Eq. 1, we can obtain the relationship between (x_F, y_F, z_F) and (u,v) as:

$$\begin{pmatrix} u \\ v \end{pmatrix} = \frac{-a}{p}\frac{1}{z_F}\begin{pmatrix} x_F \\ y_F \end{pmatrix} + \begin{pmatrix} c_u \\ c_v \end{pmatrix} \tag{12}$$

Similarly, after camera motion, let $\boldsymbol{F'}$ be imaged at pixel location $(u+\Delta u, v+\Delta v)$ which is given by:

$$\begin{pmatrix} u+\Delta u \\ v+\Delta v \end{pmatrix} = \frac{-a}{p}\frac{1}{z_F+\Delta z_F}\begin{pmatrix} x_F+\Delta x_F \\ y_F+\Delta y_F \end{pmatrix} + \begin{pmatrix} c_u \\ c_v \end{pmatrix}. \tag{13}$$

Because the motion is sufficiently small such that $z_F \gg \Delta z_F$, we use the following Taylor approximation:

$$\frac{1}{z_F+\Delta z_F} \approx \frac{1}{z_F}\left(1 - \frac{\Delta z_F}{z_F}\right). \tag{14}$$

From Eq. 12, 13 and 14, the focal speckle motion $\boldsymbol{\Delta F}$ and the speckle flow $(\Delta u, \Delta v)$ are related as:

$$\begin{pmatrix} \Delta u \\ \Delta v \end{pmatrix} \approx \frac{-a}{pz_F}\frac{1}{z_F}\begin{pmatrix} z_F\Delta x_F - x_F\Delta z_F - \Delta x_F\Delta z_F \\ z_F\Delta y_F - y_F\Delta z_F - \Delta y_F\Delta z_F \end{pmatrix}.$$

Next, we make the approximation that the term $\Delta x_F \Delta z_F \approx 0$, because the camera motion is sufficiently small. Then, the above can be written as:

$$\begin{pmatrix} \Delta u \\ \Delta v \end{pmatrix} \approx \frac{-a}{pz_F}\begin{pmatrix} 1 & 0 & -\frac{x_F}{z_F} \\ 0 & 1 & -\frac{y_F}{z_F} \end{pmatrix}\boldsymbol{\Delta F}. \tag{15}$$

Then, by substituting Eq. 11 into Eq. 15, we get:

$$\begin{aligned}\begin{pmatrix} \Delta u \\ \Delta v \end{pmatrix} &= \frac{-a}{pb}\begin{pmatrix} -1 & 0 & \hat{x_F} \\ 0 & -1 & \hat{y_F} \end{pmatrix}\boldsymbol{t_C} \\ &\quad + \frac{-a}{p}\begin{pmatrix} \hat{x_F}\hat{y_F} & -1-\hat{x_F^2} & +\hat{y_F} \\ 1+\hat{y_F^2} & -\hat{x_F}\hat{y_F} & -\hat{x_F} \end{pmatrix}\boldsymbol{\theta_C}\end{aligned} \tag{16}$$

Note that $z_F = b$, because the point $\boldsymbol{F}$ lies on the camera focus plane. When focal length is sufficiently long, we use the paraxial approximation, i.e., $|x_F|, |y_F| \ll |z_F|$. Therefore, $|\hat{x_F}|, |\hat{y_F}| \ll 1$, and $\hat{x_F^2} \approx \hat{y_F^2} \approx \hat{x_F}\hat{y_F} \approx 0$. Then we get the camera speckle flow model:

$$\begin{aligned}&\begin{pmatrix} \Delta u \\ \Delta v \end{pmatrix} \approx \\ &\frac{-a}{pb}\begin{pmatrix} -1 & 0 & \hat{x_F} \\ 0 & -1 & \hat{y_F} \end{pmatrix}\boldsymbol{t_C} + \frac{-a}{p}\begin{pmatrix} 0 & -1 & +\hat{y_F} \\ 1 & 0 & -\hat{x_F} \end{pmatrix}\boldsymbol{\theta_C}\end{aligned} \tag{17}$$

5 Speckle Flow Due To Source Motion

In this section, we derive the speckle flow due to light source motion. Following the notation from the main paper, the electric field at camera pixels $\boldsymbol{I}$ and $\boldsymbol{I'}$ before and after light source movement, respectively, can be expressed as:

$$E(\boldsymbol{I}) = c\iint_{\Omega} \alpha(\boldsymbol{S_\delta})\, e^{\frac{2\pi i}{\lambda}\Gamma(\boldsymbol{L},\boldsymbol{S_\delta},\boldsymbol{F})} d\boldsymbol{S_\delta}, \tag{18}$$

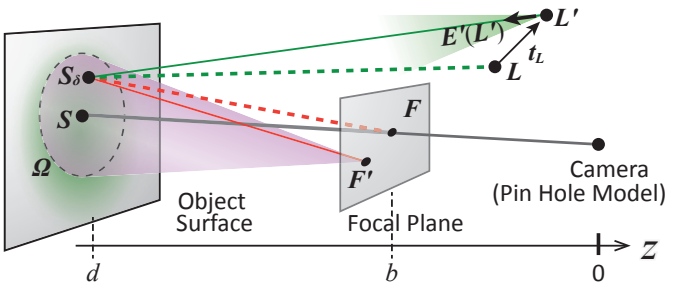


Figure 3: **Movement Of Speckle Flow Due To Light Source Motion:** The small light source movement from $\boldsymbol{L}$ to $\boldsymbol{L'}$ creates speckle flow.

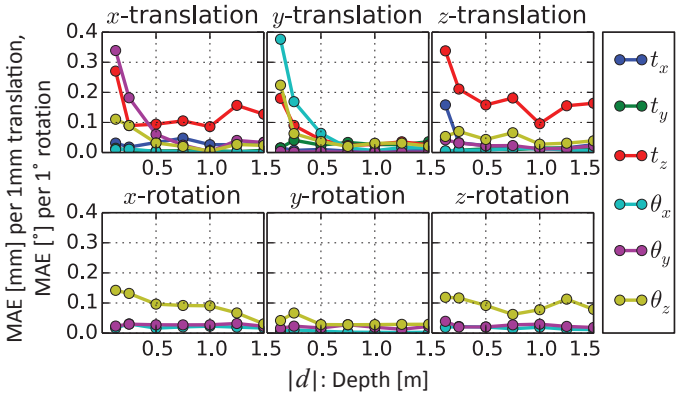


Figure 4: **Estimated Error Vs. Scene Depth.**

$$E'\left(\boldsymbol{I'}\right) = c' \iint_{\Omega'} \alpha\left(\boldsymbol{S_\delta}\right) e^{\frac{2\pi i}{\lambda}\Gamma\left(\boldsymbol{L'},\boldsymbol{S_\delta},\boldsymbol{F'}\right)} d\boldsymbol{S_\delta}, \tag{19}$$

where $\boldsymbol{S}$ is the surface point on the line joining camera center and original focus point $\boldsymbol{F}$, and $\boldsymbol{S_\delta}$ is a point in the neighborhood of $\boldsymbol{S}$ defined by the blur kernel Ω. $\boldsymbol{F'}$ is a point on the camera focus plane that is conjugate to pixel $\boldsymbol{I'}$ after light source motion. This is illustrated in Figure 3. Note that, $c \approx c' \approx \nu\frac{E(L)}{\Gamma(L,S)}$. This is because the distance between source and surface $\Gamma(L,S)$ is assumed to be significantly large, and approximately constant over the surface points.

If the optical path lengths $\Gamma\left(\boldsymbol{L},\boldsymbol{S_\delta},\boldsymbol{F}\right)$ and $\Gamma\left(\boldsymbol{L'},\boldsymbol{S_\delta},\boldsymbol{F'}\right)$ satisfy the following condition:

$$\Delta\Gamma\left(\boldsymbol{L},\boldsymbol{S_\delta},\boldsymbol{F}\right) = \Gamma\left(\boldsymbol{L'},\boldsymbol{S_\delta},\boldsymbol{F'}\right) - \Gamma\left(\boldsymbol{L},\boldsymbol{S_\delta},\boldsymbol{F}\right) = \text{Constant} \tag{20}$$

for all points $\boldsymbol{S_\delta}$, then the observed speckle brightness before and will be the same, i.e., $|E\left(\boldsymbol{I}\right)|^2 = |E'\left(\boldsymbol{I'}\right)|^2$.

Next, note that $\Gamma\left(\boldsymbol{L},\boldsymbol{S_\delta},\boldsymbol{F}\right) = |\boldsymbol{S_\delta L}| + |\boldsymbol{S_\delta F}|$, and $|\boldsymbol{S_\delta L}|$ can be approximated using Taylor expansion as:

$$\begin{aligned} |\boldsymbol{S_\delta L}| &= |\boldsymbol{SL}|\left(1 - 2\frac{\boldsymbol{SL}\cdot\boldsymbol{\delta}}{|\boldsymbol{SL}|^2} + \frac{|\boldsymbol{\delta}|^2}{|\boldsymbol{SL}|^2}\right)^{\frac{1}{2}}, \\ &\approx |\boldsymbol{SL}| - \boldsymbol{sl}\cdot\boldsymbol{\delta}, \end{aligned} \tag{21}$$

where $\boldsymbol{sl}$ is the unit vector along $\boldsymbol{SL}$ such that $\boldsymbol{sl} = \frac{\boldsymbol{SL}}{|\boldsymbol{SL}|}$. The other path lengths $|\boldsymbol{S_\delta F}|$, $|\boldsymbol{S_\delta L'}|$, $|\boldsymbol{S_\delta F'}|$ can be given in the same way. Therefore, $\Delta\Gamma\left(\boldsymbol{L},\boldsymbol{S_\delta},\boldsymbol{F}\right)$ can be expressed as:

$$\Delta\Gamma\left(\boldsymbol{L},\boldsymbol{S_\delta},\boldsymbol{F}\right) = \Delta|\boldsymbol{SF}| + \Delta|\boldsymbol{SL}| - \left(\boldsymbol{\Delta sf} + \boldsymbol{\Delta sl}\right)\cdot\boldsymbol{\delta}. \tag{22}$$

In order to satisfy Eq. 20 at all point $\boldsymbol{S_\delta}$, the following equation must hold:

$$\frac{\partial}{\partial\boldsymbol{\delta}}\Delta D\left(\boldsymbol{L},\boldsymbol{S_\delta},\boldsymbol{F}\right) = \boldsymbol{\Delta sf} + \boldsymbol{\Delta sl} = 0 \tag{23}$$

Let the vector $\Delta\boldsymbol{F} = \boldsymbol{F'} - \boldsymbol{F} = (\Delta x_F,\ \Delta y_F,\ \Delta z_F)^T$ denote the motion of point $\boldsymbol{F}$ in CCS due to light source motion. Then, from Eq. 23, the light source movement $\boldsymbol{t_L} = (\Delta x_L, \Delta y_L, \Delta z_L)^T$ and the vector $\Delta\boldsymbol{F}$ are related as:

$$\begin{aligned} &\left(\frac{\partial\boldsymbol{sf}}{\partial x_F}\Delta x_F + \frac{\partial\boldsymbol{sf}}{\partial y_F}\Delta y_F\right) \\ &+ \left(\frac{\partial\boldsymbol{sl}}{\partial x_L}\Delta x_L + \frac{\partial\boldsymbol{sl}}{\partial y_L}\Delta y_L + \frac{\partial\boldsymbol{sl}}{\partial z_L}\Delta z_L\right) = 0. \end{aligned} \tag{24}$$

We write Eq. 24 into matrix form as:

$$\mathbf{P_{SF}}\begin{pmatrix}\Delta x_F\\ \Delta y_F\end{pmatrix} = -\mathbf{P_{SL}}\begin{pmatrix}\Delta x_L\\ \Delta y_L\\ \Delta y_L\end{pmatrix}, \tag{25}$$

$$\mathbf{P_{SF}} = \frac{1}{|\boldsymbol{SF}|}\begin{pmatrix}1-x_{sf}^2 & -y_{sf}x_{sf}\\ -x_{sf}y_{sf} & 1-y_{sf}^2\\ -x_{sf}z_{sf} & -y_{sf}z_{sf}\end{pmatrix}, \tag{26}$$

$$\mathbf{P_{SL}} = \frac{1}{|\boldsymbol{SL}|}\begin{pmatrix}1-x_{sl}^2 & -y_{sl}x_{sl} & -z_{sl}x_{sl}\\ -x_{sl}y_{sl} & 1-y_{sl}^2 & -z_{sl}y_{sl}\\ -x_{sl}z_{sl} & -y_{sl}z_{sl} & 1-z_{sl}^2\end{pmatrix}. \tag{27}$$

Multiplying the pseudo-inverse matrix of $\mathbf{P_{SF}}$, the speckle flow in sensor plane is calculated as:

$$\begin{pmatrix}\Delta u\\ \Delta v\end{pmatrix} = -\frac{-a}{bp}\mathbf{P_{SF}}^{\dagger}\mathbf{P_{SL}}\boldsymbol{t_L}. \tag{28}$$

Under the paraxial approximation and assuming laser position is closer to the camera, $x_{sf}^2 \approx y_{sf}^2 \approx x_{sf}y_{sf} \approx 0$, $x_{sf}z_{sf} \approx x_{sf}$. In addition, $\frac{1}{|\boldsymbol{SF}|} \approx \frac{1}{b-d}$ and $\frac{1}{|\boldsymbol{SL}|} \approx \frac{1}{-d}$. Then, $\boldsymbol{sl}$ can be approximated as:

$$\boldsymbol{sl} = \frac{\boldsymbol{SL}}{|\boldsymbol{SL}|} \approx -\frac{1}{d}\left(\boldsymbol{L}-\boldsymbol{S}\right) = \begin{pmatrix}\hat{x_F} - \frac{1}{d}x_L\\ \hat{y_F} - \frac{1}{d}y_L\end{pmatrix}. \tag{29}$$

Note that, $\frac{\boldsymbol{S}}{d} = \frac{\boldsymbol{F}}{b} = (\hat{x_F}, \hat{y_F}, 1)^T$, because $\boldsymbol{S}$, $\boldsymbol{F}$ and the lens center are on the same line. Then, the $\mathbf{P_{SF}}^{\dagger}$ will be,

$$\mathbf{P_{SF}}^{\dagger} \approx |\boldsymbol{SF}|\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\end{pmatrix}. \tag{30}$$

From Eq. 27 to 30, the speckle flow model is obtained as:

$$\begin{pmatrix}\Delta u\\ \Delta v\end{pmatrix} = \frac{-a}{p}\left(\frac{1}{b}-\frac{1}{d}\right)\begin{pmatrix}-1 & 0 & \hat{x_F} - \frac{x_L}{d}\\ 0 & -1 & \hat{y_F} - \frac{y_L}{d}\end{pmatrix}\boldsymbol{t_L}. \tag{31}$$

6 Quasi-Invariance To Scene Depth

We conducted precise measurements of ego-motion estimation errors for different scene depth. we performed ego-motion estimation with the scene (a single fronto-parallel plane) placed

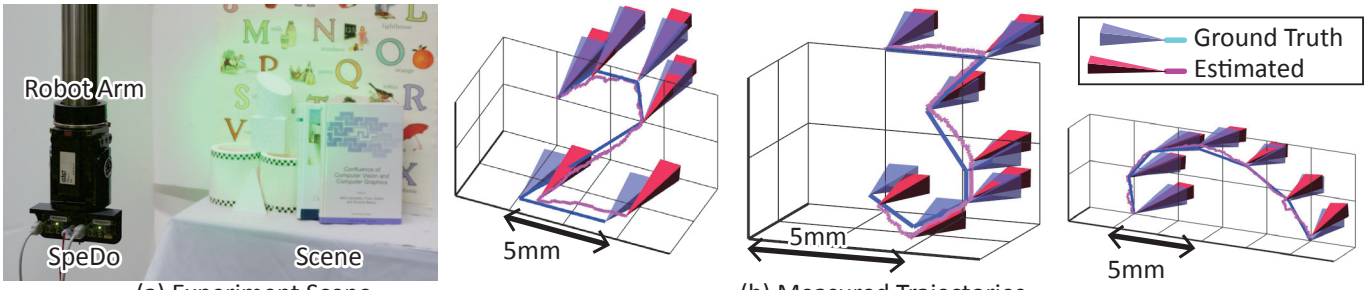


Figure 5: **More Ego-Motion Estimation Results Using Spedo For Complex Trajectories.**

at different scene depths between 1.5 meters and 0.125 meters, and measured the mean absolute error (MAE) for each scene depth.

Fig. 4 shows the example error plots for six different trajectories (translations and rotations along three axes). For each trajectory, the corresponding sub-figure shows six error plots, one each for one degree of motion (e.g. blue plots means MAE of x-translation estimation). When scene depth is larger than 0.5 meters, the error is less than 0.05 mm and 0.05°. However, errors are larger if the depth is smaller than 0.5 meters. The plots shows that z translation t_z and z-rotation θ_z have larger errors, as discussed in the main paper.

7 More Ego-Motion Estimation Results

SpeDo system can measure ego-motion on complex scenes with or without textures with a range of depths. Fig. 5 (b) shows ego-motion estimation results using SpeDo on the scene shown in Fig. 5 (a) for several trajectories, including motion along roman numerals 2 and 3. **For more and video results, please see the supplementary video.**