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

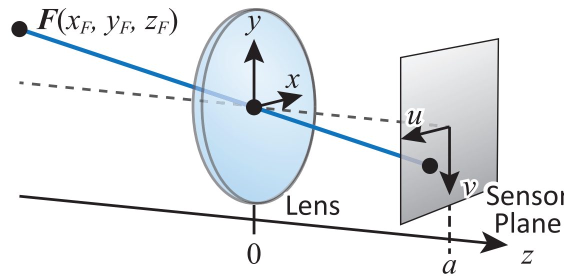

Figure 1: **Camera Coordinate System.**

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.

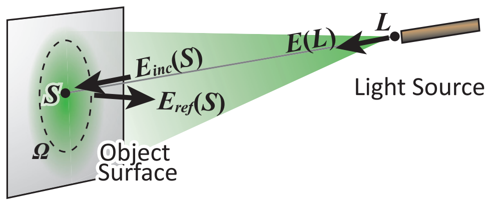

Figure 2: **The Electric Field After Reflection:** The point light source at location $\boldsymbol{L}$ emits coherent light and the object surface refrects the light. The reflected light at location $\boldsymbol{S}$ can be calculated from the distance between $\boldsymbol{L}$ and $\boldsymbol{S}$, and 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 $\boldsymbol{\Delta F} = \boldsymbol{F'} - \boldsymbol{F} = (\Delta x_F,\ \Delta y_F,\ \Delta z_F)^T$ denote the motion of point $\boldsymbol{F}$ in CCS due to camera motion:

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

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:

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

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:

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

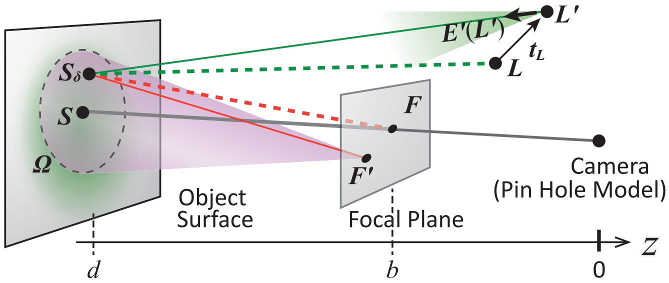

Figure 3: **Movement Of Speckle Flow Due To Light Source Motion:** The small light source movement from $\boldsymbol{L}$ to $\boldsymbol{L'}$ creates speckle flow.

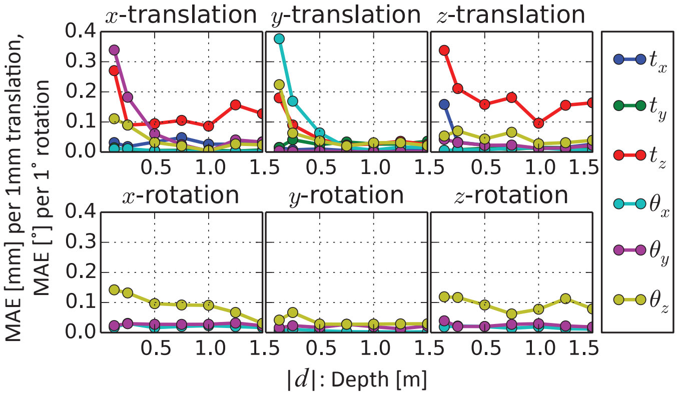

Figure 4: **Estimated Error Vs. Scene Depth.**

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

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:

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

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:

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

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 $\boldsymbol{\Delta F} = \boldsymbol{F'} - \boldsymbol{F} = (\Delta x_F,\ \Delta y_F,\ \Delta z_F)^T$ 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 $\boldsymbol{\Delta F}$ 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

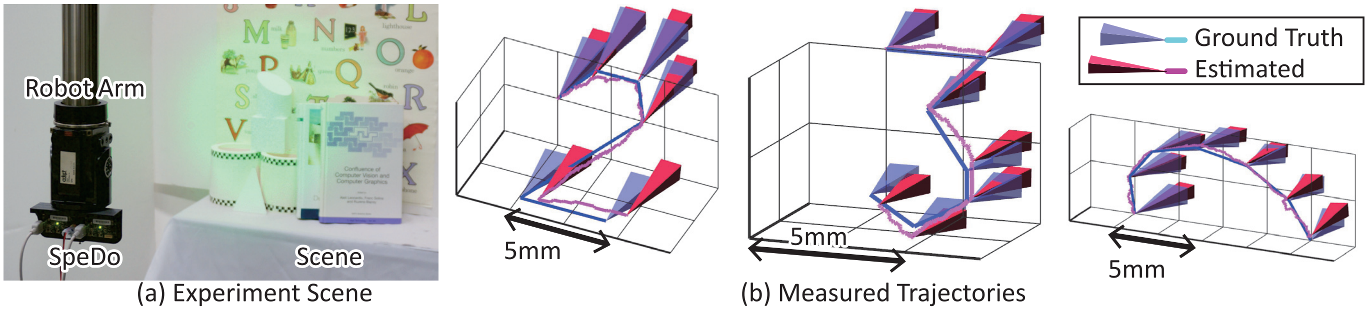

Figure 5: **More Ego-Motion Estimation Results Using Spedo For Complex Trajectories.**

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.**