Supplementary material: Robust Frame-to-Frame Camera Rotation Estimation in Crowded Scenes

These supplementary materials include the derivation to find the point of intersection between the lines of compatible rotations and the plane z=0 in rotation space, as well as detailed evaluations and descriptions of our dataset. In addition to this document, we include a video of a rotation stabilization application using our method against the best of the baselines, and the MATLAB implementation of our method which is located in the folder *code/*.

Appendix **A** shows additional results. In Appendix **B** we show additional derivations on the line of compatible rotations using Longuet-Higgins motion model. In Appendix **C** we evaluate the effect of residual translational flow on the rotation estimation error. We conclude these supplementary materials in Appendix **D** with a detailed assessment of our new BUSS dataset.

A. Additional Results

We show some qualitative results (see the video attached) and frame-per-frame results in a complete sequence.

A.1. Qualitative results

One way to visually assess the quality of the rotation estimations is to use the estimate of the rotations to remove the camera rotations from videos. More specifically, this can be done by first taking the inverse of the rotation estimate, then generating the pure rotational optical flow corresponding to this rotation, and finally warping the frame of the video using the flow. A perfect estimate of the rotation would entirely remove the camera's rotational motion stabilizing distant points.

We include the qualitative result for one video from our BUSS dataset in the folder *qualitative_results/results.mp4*. First we show the video without stabilization for $\frac{1}{3}$ of the frames. Next, we stabilize the video with our method and Bruss & Horn's method each for $\frac{1}{3}$ of the frames. Finally, we show a collage of the entire video with no stabilization, stabilized video using the ground truth rotation, our estimate, and Bruss & Horn's estimate. For our estimate, we see almost no motion when examining the distant objects in the scene (such as the buildings) after stabilizing the video using our rotation estimate. That is because distant objects are not affected by translation; hence, removing the rotation stabilizes the objects. On the other hand, objects close to the camera will be affected by camera translation and therefore are not stabilized.

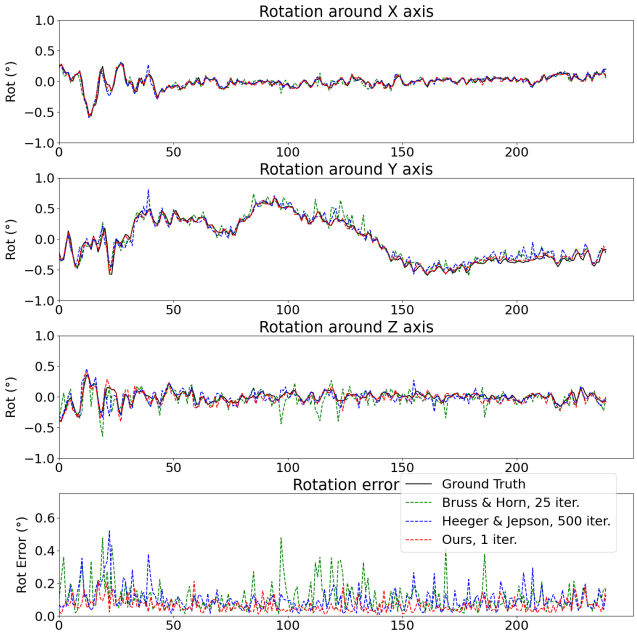

Figure A.1. **Comparison of selected methods on one video from the BUSS dataset**. The three upper graphs show the x, y and z components of the rotations per frame estimated by ours, Bruss & Horn, and Heeger & Jepson as well as the ground truth. The lower graph shows the rotation error of the three methods.

A.2. Detailed results

We provide the frame-per-frame results of the three more competitive methods in Fig. **A.1**. We show the three rotation angles (rotation around x, y, and z-axis) for ground truth, Bruss & Horn, Heeger & Jepson, and our method. For all axis of rotation (top 3 graphs), the prediction of our method follows the ground truth without any significant deviations. In contrast, the other methods have large variations visible as spikes on the graph. In the last row of Fig. **A.1** we show the rotation error per frame. Our method has consistently

smaller rotational errors than other methods.

B. Line of Compatible Rotations

In this section, we show that the lines of compatible rotations for the Longuet-Higgins motion model, which is expressed as

$$\mathbf{v} = \underbrace{\begin{pmatrix} \frac{A}{f}xy - Bf - \frac{B}{f}x^2 + Cy \\ Af + \frac{A}{f}y^2 - \frac{B}{f}xy - Cx \end{pmatrix}}_{\mathbf{v_r}} + \underbrace{\begin{pmatrix} \frac{-fU+xW}{Z} \\ \frac{-fV+yW}{Z} \end{pmatrix}}_{\mathbf{v_t}}, \quad (1)$$

always intersect the $Z = 0$ plane.

As explained in Section 3.1.2 of the paper, the vector that gives the direction of the line l of compatible rotations is $\mathbf{d} = \mathbf{n_u} \times \mathbf{n_v}$, where $\mathbf{n_u}$ and $\mathbf{n_v}$ are the normal vectors to the planes defined by $\mathbf{v_r}$.

Taking the cross product gives:

$$\mathbf{d} = \mathbf{n_u} \times \mathbf{n_v} = \begin{bmatrix} \frac{xy^2}{f} + \frac{x(f^2+x^2)}{f} \\ \frac{y(f^2+y^2)}{f} + \frac{x^2y}{f} \\ \frac{(f^2+x^2)(f^2+y^2)}{f^2} - \frac{x^2y^2}{f^2} \end{bmatrix} \quad (2)$$

$$= \underbrace{(y^2 + f^2 + x^2)}_{>0} \begin{bmatrix} \frac{x}{f} \\ \frac{y}{f} \\ 1 \end{bmatrix} \quad (3)$$

The z component of $\mathbf{d}$ cannot be 0 since it would imply a focal length of 0. Therefore, the line l can't be co-planar to the plane $C = 0$, and thus we can complete the definition of l by finding its intersection with the plane $C = 0$, and solving the system:

$$\mathbf{v} = \begin{pmatrix} \frac{A}{f}xy - Bf - \frac{B}{f}x^2 \\ Af + \frac{A}{f}y^2 - \frac{B}{f}xy \end{pmatrix} \quad (4)$$

C. Rotation estimation error due to camera translation

Figure C.2 shows the rotation estimation error of our method given pure translation flows. We generated pure translational X, Y and Z flows for different translation speed and depths, and use our method to estimate the rotation, which should be zero. The corruption of the flow by the translation depends on many factors, such as the translation speed and direction, focal length, and scene geometry. A pure z translation at 1 m/s, with points in the scene more than 25 meters away will have no impact on rotation error. For pure x and y translations (respectively confusable with y and x rotations), the rotation estimate is only off by 1 bin (0.057 deg) for depths between 25 and 60 meters and 0 bins for distances greater than 60 m.

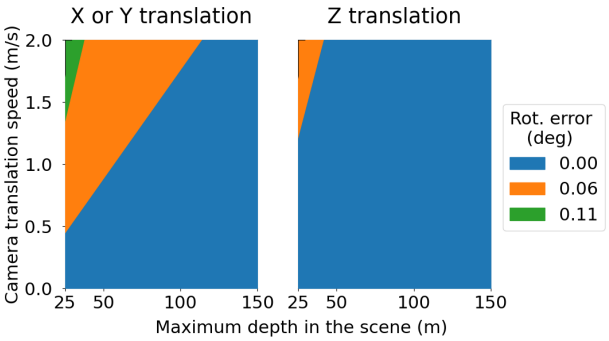

Figure C.2. Rotation estimation error in the pure translation case with respect to the camera translation speed and the maximum depth in the scene. Left: Rotation error for pure X or Y translation motions. Right: Rotation error for pure Z translation motions.

D. BUSS's ground truth accuracy assessment

We provide additional details and results about the ground truth accuracy evaluation experiment. As mentioned in Sec. 4.1 of the paper, we compared the gyroscope readings of the OPPO A5 2020 (used for the dataset) with the gyroscope of an iPhone 12 mini. Figure **D.3** shows the setup we used to rigidly mount the two phones. Since the two gyroscopes have different frequencies (OPPO has 400Hz and iPhone has 100Hz), we compare their measurements by linearly interpolating the OPPO's gyroscope readings. In other words, we want to get readings from the OPPO's gyroscope at the same timestamps as the iPhone's gyroscope readings.

Figure D.3. **Set up to test the accuracy of the gyroscope**. Rig to ensure identical rotations between the two phones to assess the accuracy of the ground truth of the BUSS dataset.

Figure **D.4** shows the histogram of the angular velocity errors between the two phones, and Figure **D.5** shows the histogram of the rotation error between the two phones over periods of 1/30 seconds. These graphs show a strong agreement between the two sensors. It is highly unlikely for the two sensor to agree if their measurements were incorrect. Therefore, this strongly suggest that the gyroscope measurements is a good ground truth for frame-to-frame rotation estimation.

As explained in Sec. 4.1 of the paper, the two phones

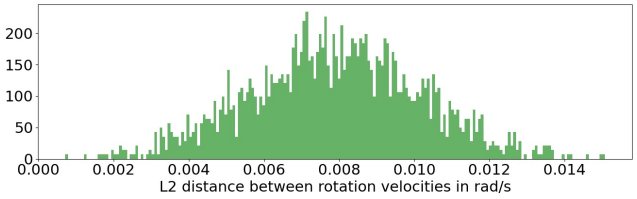

Figure D.4. **Comparison of the rotation velocities between the two phones**. Histogram of the L^2 distance between the rotation velocities of the two phones

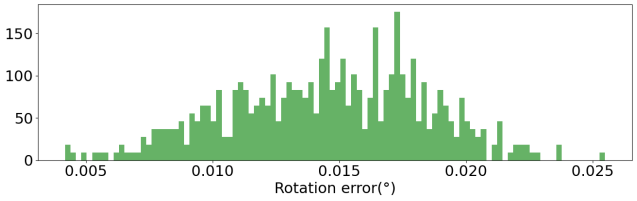

Figure D.5. **Comparison of the rotation error between the two phones**. Histogram of the amplitude of the rotation errors between the frame-to-frame (30fps) rotation estimation of the two gyroscopes.

need to be aligned temporally and spatially. In this section, we explain how we corrected those two types of misalignment. We confirm gyroscope agreement in Figure D.6.

Temporal misalignment: Each gyroscope measurement is associated with a timestamp referencing the internal clock of each phone. To compare the gyroscope measurements across the two phones, we need to synchronize their clocks. To do so, we searched for the shift (at the precision of 0.01 ms) that minimizes the mean square error between the rotation velocity measures of the two gyroscopes.

Spatial misalignment: Another source of misalignments between the two gyroscopes is the relative orientation R between the two sensors. More concretely, we want to express the angular rotations measured by the OPPO phone $\omega_{\text{oppo[oppo]}}$ in the frame of reference of the iPhone $\omega_{\text{oppo[iphone]}}$ where $\omega_{\text{oppo[oppo]}}R = \omega_{\text{oppo[iphone]}}$. We estimate R by minimizing the mean squared errors between $\omega_{oppo[oppo]}R$ and $\omega_{oppo[iphone]}$ using the Kabsch algorithm.

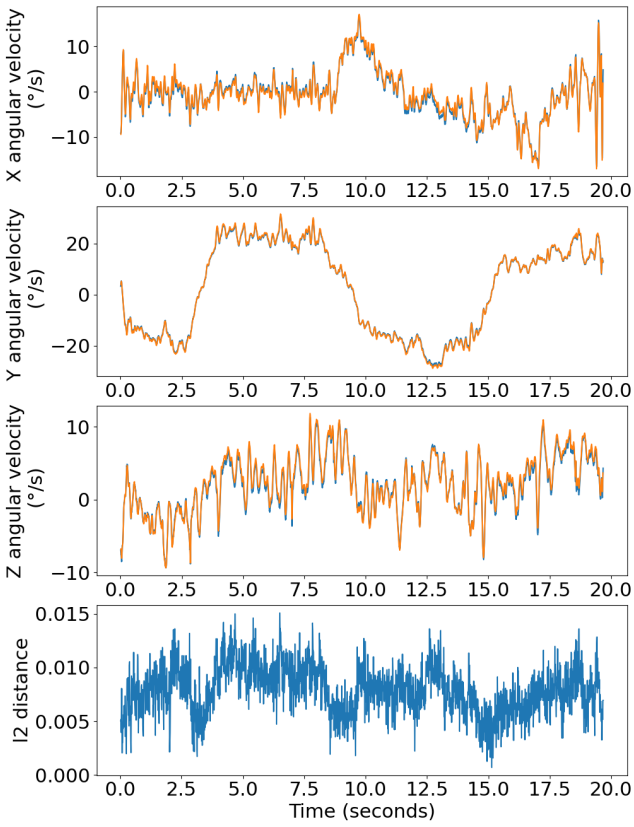

Figure D.6. **Comparison of the gyroscope measurements from the two phones**. The three upper graphs show the x, y and z rotation velocities of the two phones. There are two curves (an orange and a blue one), but they are on top of each other. The lower graph shows the error between the two gyroscopes.