Shape from Depth Discontinuities under Orthographic Projection
Supplementary Material

1. Supplementary Experimental Results

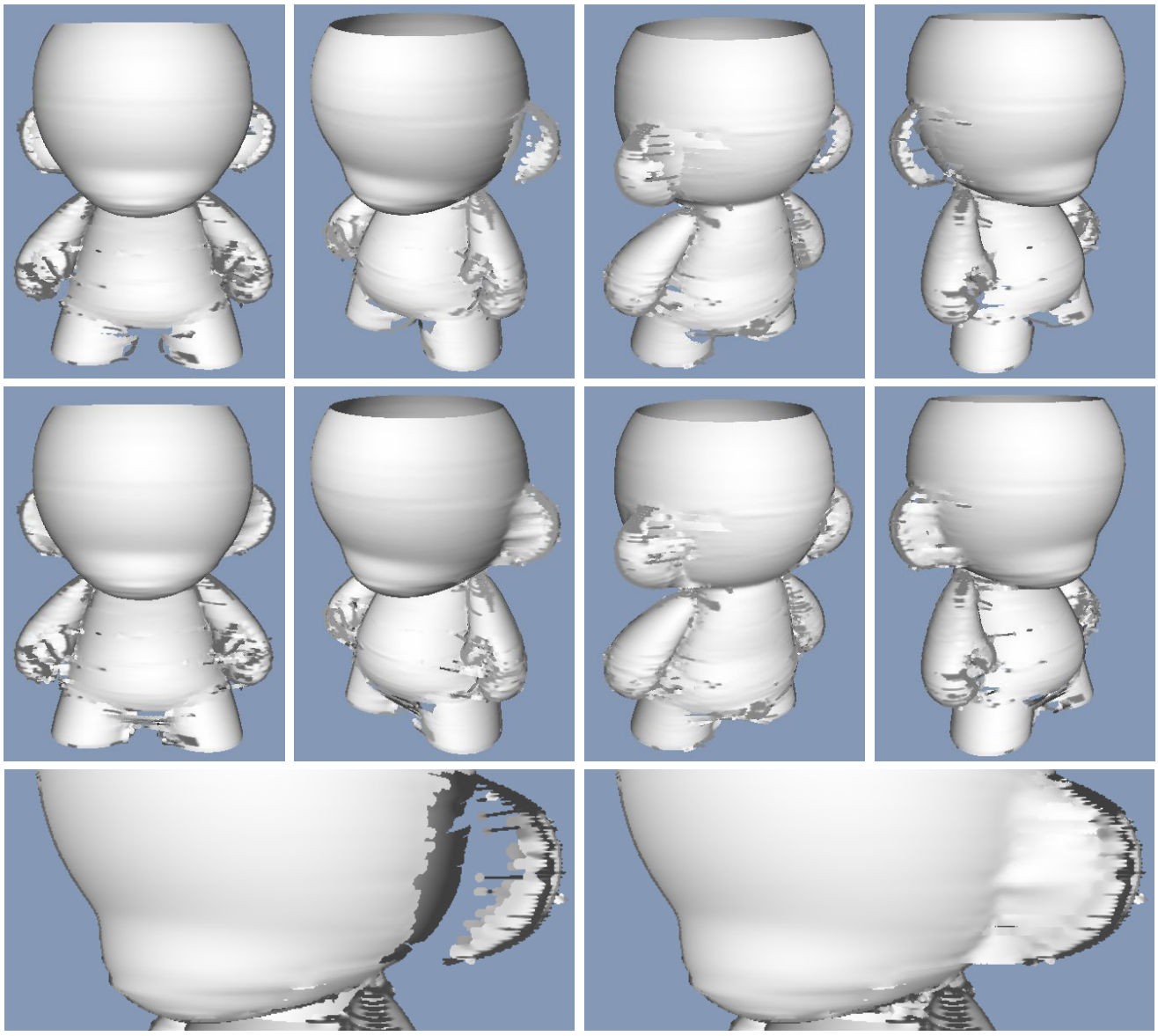

Figure 1. Comparison of 3-D surface reconstructions. Each epipolar-plane depth edge confidence image, as shown in supplementary Figure 2, was reconstructed independently using the method outlined in Sections 4.1 and 4.2 of the paper. The gaps in the reconstruction were filled using the procedure described in Section 4.3. All renderings were produced using Pointshop3D, with surface normal estimates provided using the Diffuser plug-in. Shading variation is due to surface normal estimation errors using the plug-in. (First row) Four views of the recovered 3-D point cloud. Note the gaps near the ears, hands, and feet. (Second row) Reconstructed point clouds with gaps filled using the proposed method. (Third row, left) Magnified view of gap near the ear. (Third row, right) Magnified view of gap completion. Note that the independent surface completions exhibit strong coherent across neighboring layers. While additional post-processing of the point cloud was not applied in this example, such processing would improve inter-layer coherence.

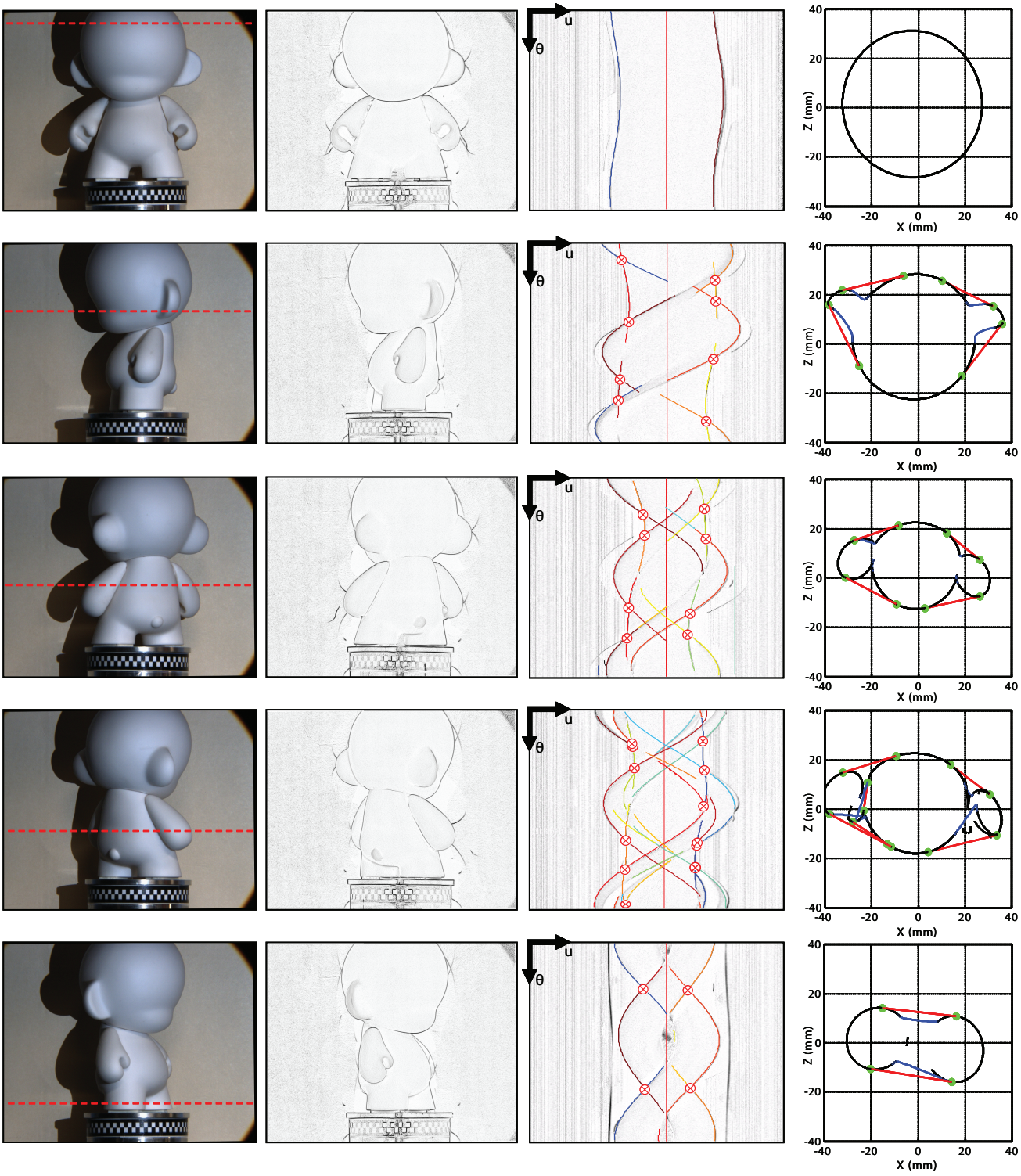

Figure 2. Orthographic multi-flash 3-D photography. (First column) Five input views, equally-spaced by a rotation of 72 degrees, acquired with a flash located to the right of the scene. (Second column) Corresponding depth edge confidence images, estimated using the procedure described in Section 4.1. (Third column) Epipolar-plane images, extracted along the corresponding dashed red lines shown in the first column (*i.e.*, from top to bottom, parallel slices through the head, ears, shoulders, hands, and feet). Tracked contours, using the procedure from Section 4.2, are shown as separately colored lines. T-junctions corresponding to local concavities are indicated with red crosses. (Fourth column) Recovered external boundaries, shown in black, for corresponding image rows. Points on either side of each concavity, corresponding to EPI T-junctions, are linked with red lines. The piecewise cubic Hermite polynomial completions are shown in blue.

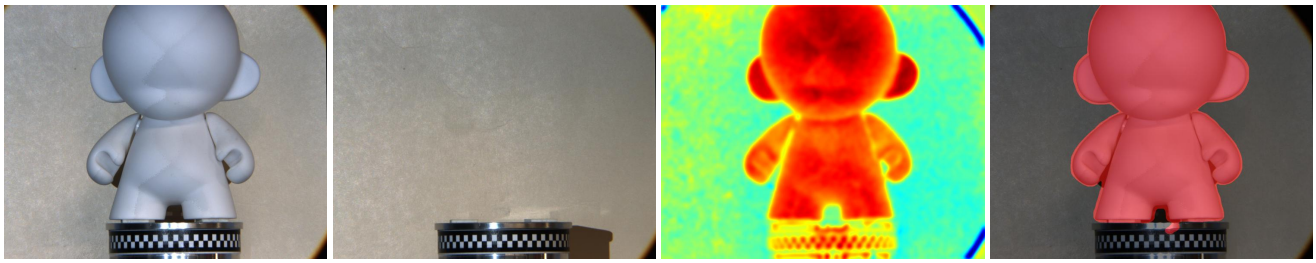

Figure 3. Chrominance-domain foreground segmentation. Automatic segmentation of the foreground object from the background is accomplished using simple background subtraction. From left to right, (1) the maximum flash composite obtained by storing the per-pixel maximum luminance sample observed over the multi-flash sequence, (2) the selected background image using a single flash located to the left of the scene, (3) absolute difference in Q-chrominance values, calculated in the YIQ color space, after background subtraction, rendered using the jet color map in Matlab so that large differences are red and small differences are green, and (4) final foreground segmentation, composited in red, after thresholding and applying a morphological closing operation to detected foreground pixels. A set of 25 foreground segmentations, evenly-spaced within the turntable sequence, is used to eliminate any tracked contours (*e.g.*, those shown in supplementary Figure 2) or reconstructed points (*e.g.*, those shown in supplementary Figure 1) that are located outside the visual hull.

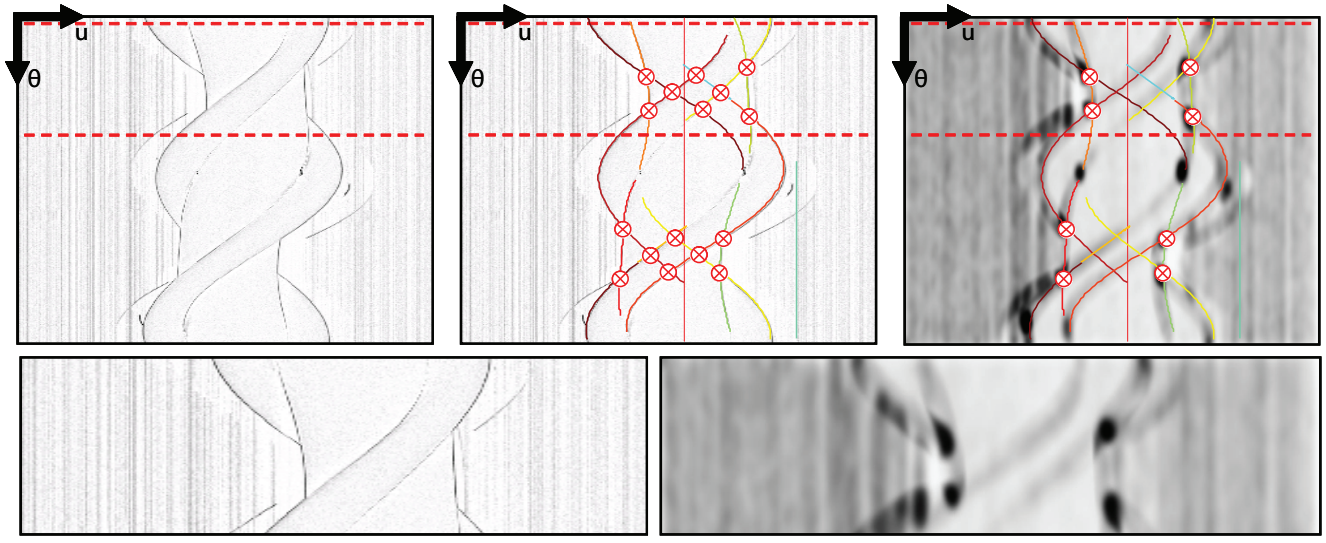

Figure 4. Epipolar-plane ridge and T-junction detection and filtering. As explained in Section 4.2 of the paper, epipolar-plane image analysis is required to: (1) track and link ridges in each EPI, (2) detect T-junctions corresponding to local surface concavities, and (3) find corresponding endpoints of each tracked contour meeting at a T-junction. (Top row, left) A single epipolar-plane depth edge confidence image, identical to that shown in the third row of supplementary Figure 2. (Top row, middle) Detected EPI ridges, modeled as trigonometric polynomials, after merging all segments given by Equations (4) and (5) in the paper. All pairwise-intersections of contours are indicated using red crosses. (Top row, right) Tracked contours and T-junctions after filtering using the T-junction likelihood map, estimating using the Harris corner measure calculated using a pair of steerable filters. Note that only intersections corresponding to a pair of visible contours are retained. (Bottom row, left) Inset region of epipolar plane image, within the dashed red lines shown in the top row. (Bottom row, right) Inset region of Harris corner measure, with darker regions corresponding to higher corner likelihood.

2. Supplementary Analysis of Orthographic Calibration

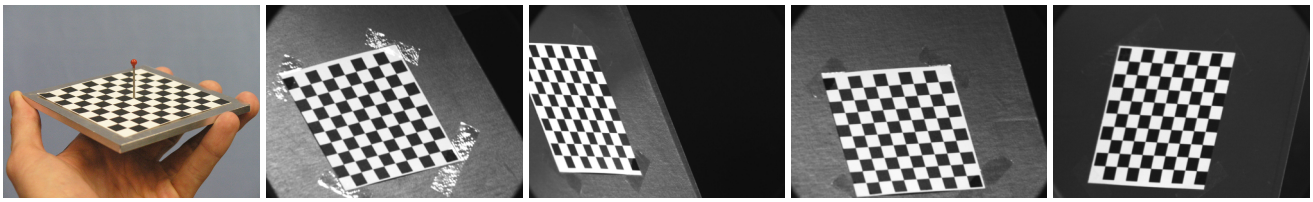

Figure 5. (Left) The red pole tip disambiguates between equivalent orientations described in Section 5.2 of the paper. (Right) Four orthographic calibration images, taken from a set of 100. Note that the calibration pole is optional for intrinsic calibration. The checkerboard corners were automatically detected using OpenCV. The methods from Sections 5.1 and 5.3 were used to estimate the intrinsic parameters.

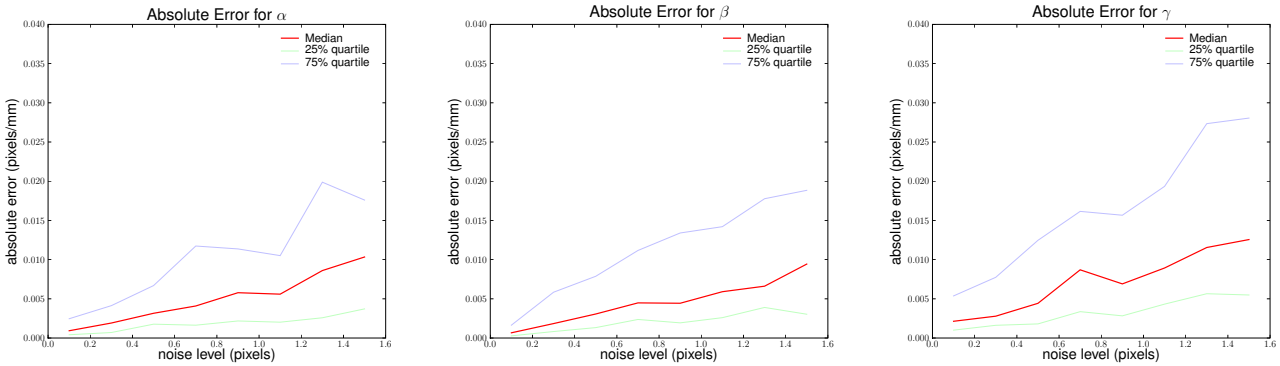

Figure 6. Sensitivity of intrinsic parameter estimation to checkerboard corner localization errors. A simple simulation is used to assess the sensitivity of the intrinsic calibration procedure described in Sections 5.1 and 5.3. For each simulated checkerboard position and pose, a checkerboard containing 10×10 squares of side length 8.5mm is projected into a synthetic image plane. The projection model corresponds to that given by Equation (2), where the intrinsic parameters are held constant with the scale factors $\{\alpha = 9.5, \beta = 10.5\}$ and skew factor $\gamma = 0.1$. The extrinsic parameters of the checkerboard pattern are randomly generated, beginning with a checkerboard centered at the world coordinate origin and containing the plane $Z = 0$. A random pose is selected as follows: (1) a rotation is applied around the X axis, by an angle uniformly-distributed on $[0, \frac{\pi}{4}]$, (2) a rotation is applied around the Z axis, by an angle uniformly-distributed on $[0, 2\pi]$. Afterwards, the position of the checkerboard is randomly translated in the XY plane, where the orientation of the translation vector is chosen from $[0, 2\pi]$. The magnitude of the translation is distributed uniformly on the interval $[0, s]$, where s represents the checkerboard side length. Errors in the corner localization procedure are simulated by adding zero-mean Gaussian noise $\sim N(0, \sigma)$ to the projected checkerboard corners. For each value of σ, 100 trials are performed using 4 simulated calibration images. The absolute estimation errors, for each of the three intrinsic parameters, are shown above as a function of the standard deviation of the additive noise process. The red curves denote the median of the absolute errors, while the 25^{th} and 75^{th} percentiles are shown in green and blue, respectively. Note that the absolute errors in the estimates increase linearly with the standard deviation of the noise process.

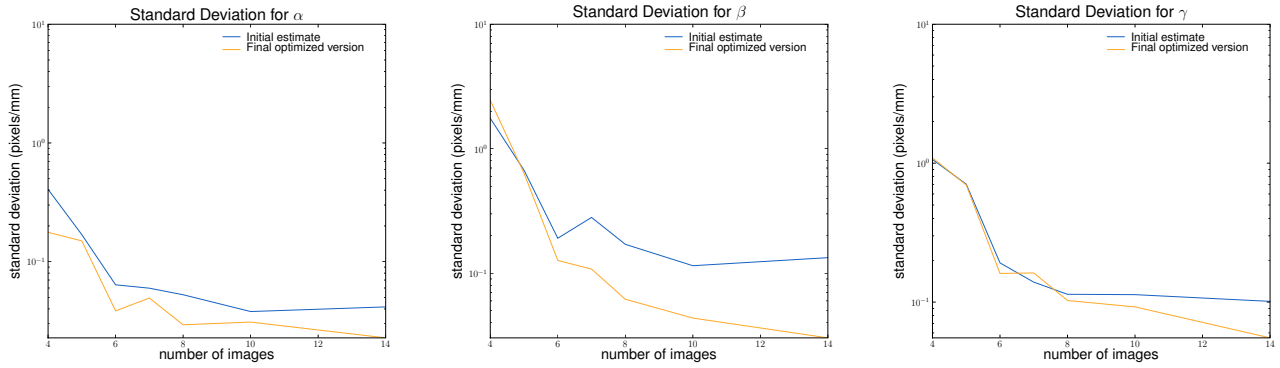

Figure 7. Sensitivity of intrinsic parameter estimation to the number of calibration images. A set of 100 calibration images, similar to supplementary Figure 5, were recorded using the orthographic camera described in Section 6 and with the dimensions given in supplementary Figure 6. For each trial a random subset of $n_{\mathcal{I}}$ calibration images were used to estimate the intrinsic parameters, including the scale factors $\{\alpha, \beta\}$ and skew factor γ. For each value of $n_{\mathcal{I}} = \{4, 5, 6, 7, 8, 10, 14\}$ 100 trials were used to estimate the standard deviation for each intrinsic parameter estimate (with the exception of $n_{\mathcal{I}} = 14$, where 60 estimates were used). The standard deviations for each parameter, as a function of the number of calibration images, are shown above. The results using only the initial estimates given by Section 5.1 and the optimized estimates given by Section 5.3 are denoted in blue and red, respectively. We observe the unexpected result that the standard deviation of β is *larger* after optimization for $n_{\mathcal{I}} = 4$. This result is likely due to multiple local minima being positioned close to the true value of β. Furthermore, a similar result is not observed for the other scale factor α; this can be attributed to the fact that α depends on a separate set of terms provided by the homographies estimated from each of the four calibration images, terms which are likely less sensitive to corner localization errors. Finally, for all three intrinsic parameters, we report that the standard deviations of their estimates converge once six or more calibration images are used.