Supplementary Material for:
Learning Object Depth from Camera Motion and Video Object Segmentation

Brent A. Griffin and Jason J. Corso
University of Michigan
{griffb,jjcorso}@umich.edu

Supplementary Material

Least-squares Solution for Object Depth

In previous work [1], we propose a least-squares object depth solution (VOS-DE) that uses more than two observations to add robustness for camera position and segmentation errors. We include this solution here for reference. The VOS-DE formulation derives an alternative form of (7) from our current paper as
\[
\hat{z}_{\text{object}} \sqrt{a_i} + c = z_i \sqrt{a_i},
\]
which over \( n \) observations in \( Ax = b \) form yields
\[
\begin{bmatrix}
\sqrt{a_1} & 1 \\
\sqrt{a_2} & 1 \\
\vdots & \vdots \\
\sqrt{a_n} & 1
\end{bmatrix}
\begin{bmatrix}
\hat{z}_{\text{object}} \\
\hat{c}
\end{bmatrix} =
\begin{bmatrix}
z_1 \sqrt{a_1} \\
z_2 \sqrt{a_2} \\
\vdots \\
z_n \sqrt{a_n}
\end{bmatrix}.
\]
Solving (2) for \( \hat{z}_{\text{object}} \) does provide a more robust depth estimate than the two-observation solution (8) in our current paper. However, our learning-based approach from Section 4 outperforms both analytic solutions in experiments.
ODMS Random Object Mask Examples

We provide a few random object mask examples using ODMS’s data-generation framework from Section 5.1 of the paper. These synthetic object examples are shown in Fig. 1 and demonstrate the Bézier curve behaviors associated with changing parameters $r_B$ and $\rho_B$.

Fig. 1. ODMS Random Object Mask Examples. All examples use $s_p = 400$, $n_p = 5$, and $\ell = \ell_{\text{min}} = 1$. $r_B$ values are 0.01, 0.05, 0.2, and 0.5 (from left to right) and $\rho_B$ values are 0.01, 0.05, and 0.2 (from top to bottom). Each generated object is unique.
ODMS Validation Results

As mentioned in Section 6 of the paper, the number of network training iterations is determined by the best validation performance, which we check at every ten training iterations. In Table 1, we provide the ODMS validation results and corresponding number of training iterations for all configurations from the paper. In general, the relative performance of each configuration is consistent between the ODMS validation and test sets.

| Config. ID | Mean Percent Error (Validation/Test) | Training Iterations |
|-----------|-------------------------------------|---------------------|
|           | Robot | Driving | Normal | Perturb | Robot | Driving | Normal | Perturb |
| Standard Configuration |         |                     |       |         |        |         |       |         |
| ODNℓ | 21.6/19.3 | 29.4/30.1 | 8.2/8.3 | 18.4/18.2 | 2390 | 1920 | 3370 | 4870 |
| ODNd | 19.6/18.5 | 32.0/30.9 | 7.9/8.2 | 18.4/18.5 | 4140 | 2990 | 3690 | 3530 |
| ODNp | 19.9/18.1 | 48.1/47.5 | 4.9/5.1 | 11.5/11.2 | 2380 | 1650 | 4740 | 4430 |
| VOS-DE | 27.4/32.6 | 35.9/36.0 | 7.9/7.9 | 34.1/33.6 | N/A | N/A | N/A | N/A |
| n = 5 Observations |         |                     |       |         |        |         |       |         |
| ODNℓ | 23.4/20.5 | 31.5/30.5 | 8.4/8.6 | 20.2/20.4 | 4100 | 1850 | 4870 | 4520 |
| ODNd | 22.8/19.5 | 34.2/31.1 | 8.4/8.4 | 20.5/20.6 | 1510 | 3450 | 3770 | 4330 |
| ODNp | 21.0/19.4 | 44.6/44.2 | 5.4/5.5 | 13.4/12.9 | 4690 | 4260 | 4980 | 4970 |
| VOS-DE | 29.5/35.1 | 34.8/34.6 | 7.8/7.9 | 32.8/32.6 | N/A | N/A | N/A | N/A |
| n = 3 Observations |         |                     |       |         |        |         |       |         |
| ODNℓ | 20.3/18.6 | 31.8/31.1 | 8.4/8.4 | 21.9/21.6 | 1820 | 2750 | 4890 | 4380 |
| ODNd | 19.9/20.6 | 34.7/33.1 | 8.4/8.4 | 21.6/21.5 | 4130 | 4320 | 4620 | 4250 |
| ODNp | 24.0/21.8 | 45.1/44.5 | 5.4/5.6 | 13.8/12.9 | 4800 | 3040 | 4990 | 4680 |
| VOS-DE | 29.5/41.2 | 45.2/44.0 | 8.0/8.1 | 37.0/35.7 | N/A | N/A | N/A | N/A |
| n = 2 Observations |         |                     |       |         |        |         |       |         |
| ODNℓ | 21.3/19.2 | 30.4/31.4 | 8.7/8.9 | 22.0/22.0 | 1140 | 1010 | 3910 | 4300 |
| ODNd | 29.1/24.2 | 39.6/35.9 | 8.6/8.9 | 21.8/21.8 | 3410 | 4570 | 3370 | 4620 |
| ODNp | 23.3/21.1 | 45.3/44.8 | 5.8/6.0 | 14.9/14.4 | 2850 | 4120 | 4610 | 4970 |
| VOS-DE | 95.8/65.5 | 55.0/41.1 | 8.2/8.3 | 90.6/86.2 | N/A | N/A | N/A | N/A |
| Perturb Training Data |         |                     |       |         |        |         |       |         |
| ODNℓ | 21.4/22.2 | 28.6/29.0 | 10.7/11.1 | 12.8/13.9 | 160 | 140 | 5000 | 5000 |
| ODNd | 25.6/25.8 | 31.4/31.4 | 11.0/11.1 | 13.1/13.2 | 420 | 2760 | 2730 | 4270 |
| ODNp | 20.5/20.1 | 59.4/69.9 | 7.0/7.3 | 8.1/8.2 | 50 | 330 | 4860 | 4780 |
| Radial Input Image |         |                     |       |         |        |         |       |         |
| ODNℓ | 13.8/13.1 | 31.6/31.7 | 8.4/8.6 | 18.2/17.9 | 1710 | 870 | 4940 | 3940 |
| ODNd | 16.6/15.2 | 30.7/30.9 | 8.3/8.4 | 18.6/18.5 | 2010 | 4200 | 4990 | 4440 |
| ODNp | 14.1/13.4 | 49.0/48.6 | 5.5/5.6 | 11.7/11.2 | 2210 | 400 | 4870 | 4710 |
ODMS Absolute Error Results

In Table 2, we provide ODMS test results for the mean absolute error, which is calculated for each example as

$$\text{Absolute Error} = |d_1 - \hat{d}_1|,$$  \hspace{1em} (3)

where $d_1$ and $\hat{d}_1$ are ground truth and predicted object depth at final pose $z_1$. Notably, our motivation to use percent error (21) in the paper is to provide a consistent comparison across domains with markedly different object depth distances. For example, the 6 cm absolute error from Fig. 7 of the paper is much better for the driving domain than it would be for robot grasping.

Table 2. Complete ODMS Validation and Test Set Results (Absolute Error)

| Config. ID | Mean Absolute Error (Validation/Test) | Training Iterations |
|------------|--------------------------------------|---------------------|
|            | Robot (cm) | Driving (m) | Normal (cm) | Perturb (cm) | Robot | Driving | Normal | Perturb |
| Standard Configuration | | | | | | |
| ODN | 7.2/6.6 | 3.8/4.3 | 3.4/4.4 | 7.3/7.2 | 2390 | 1920 | 3370 | 4870 |
| ODN | 6.4/6.0 | 4.1/4.4 | 3.1/3.1 | 7.4/7.3 | 4140 | 2990 | 3690 | 3530 |
| ODN | 6.8/6.3 | 7.1/7.8 | **1.8/1.8** | 3.9/3.7 | 2380 | 1650 | 4740 | 4430 |
| VOS-DE | 8.8/10.0 | 5.0/5.4 | 2.8/2.8 | 15.3/14.9 | N/A | N/A | N/A | N/A |

$n = 5$ Observations

| Config. ID | Mean Absolute Error (Validation/Test) | Training Iterations |
|------------|--------------------------------------|---------------------|
| ODN | 7.8/7.0 | 3.9/4.3 | 3.4/4.5 | 8.2/8.3 | 1000 | 1850 | 4870 | 4520 |
| ODN | 7.0/6.1 | 4.4/4.6 | 3.3/3.3 | 8.1/8.0 | 1510 | 3450 | 3770 | 4330 |
| ODN | 7.3/6.8 | 6.5/7.2 | 1.9/2.0 | 4.9/4.7 | 4690 | 4260 | 4980 | 4970 |
| VOS-DE | 9.6/10.8 | 5.0/5.2 | 2.9/2.9 | 14.3/14.2 | N/A | N/A | N/A | N/A |

$n = 3$ Observations

| Config. ID | Mean Absolute Error (Validation/Test) | Training Iterations |
|------------|--------------------------------------|---------------------|
| ODN | 6.8/6.3 | 4.1/4.5 | 3.4/4.4 | 8.8/8.6 | 1820 | 2750 | 4890 | 4380 |
| ODN | 6.8/7.0 | 4.4/4.7 | 3.3/3.3 | 8.6/8.4 | 4130 | 4320 | 4620 | 4250 |
| ODN | 7.9/7.3 | 6.6/7.3 | 1.9/1.9 | 4.8/4.4 | 4800 | 3040 | 4990 | 4680 |
| VOS-DE | 11.2/12.6 | 6.3/5.9 | 2.9/2.9 | 14.7/14.2 | N/A | N/A | N/A | N/A |

$n = 2$ Observations

| Config. ID | Mean Absolute Error (Validation/Test) | Training Iterations |
|------------|--------------------------------------|---------------------|
| ODN | 7.0/6.4 | 3.7/4.3 | 3.5/3.6 | 8.5/8.4 | 1140 | 1010 | 3910 | 4300 |
| ODN | 9.2/7.8 | 4.8/5.0 | 3.5/3.5 | 8.6/8.4 | 3410 | 4570 | 3370 | 4620 |
| ODN | 8.0/7.2 | 6.8/7.5 | 2.0/2.1 | 5.4/5.1 | 2850 | 4120 | 4610 | 4970 |
| VOS-DE | 36.2/21.9 | 8.5/6.7 | 3.0/3.0 | 41.1/49.7 | N/A | N/A | N/A | N/A |

Perturb Training Data

| Config. ID | Mean Absolute Error (Validation/Test) | Training Iterations |
|------------|--------------------------------------|---------------------|
| ODN | 7.0/6.9 | 3.5/4.1 | 4.3/4.5 | 5.2/5.2 | 100 | 140 | 5000 | 5000 |
| ODN | 8.4/8.5 | 4.0/4.4 | 4.4/4.4 | 5.2/5.1 | 420 | 2760 | 2730 | 4270 |
| ODN | 6.7/5.8 | 8.9/9.9 | 2.4/2.5 | **2.8/2.8** | 50 | 330 | 4860 | 4780 |

Radial Input Image

| Config. ID | Mean Absolute Error (Validation/Test) | Training Iterations |
|------------|--------------------------------------|---------------------|
| ODN | **4.4/4.3** | 4.0/4.5 | 3.5/3.5 | 7.4/7.2 | 1710 | 870 | 4940 | 3940 |
| ODN | 5.6/5.0 | 3.8/4.3 | 3.3/3.4 | 7.5/7.4 | 2010 | 4200 | 4980 | 4440 |
| ODN | 4.4/4.4 | 7.2/8.0 | 1.9/1.9 | 4.3/4.0 | 2210 | 460 | 4870 | 4710 |
ODMS Robot Test Set Segmentation Examples

For the ODMS Robot test set, we intentionally choose challenging objects, spanning from a single die to the 470 mm long pan. Not surprising, segmenting diverse objects presents varied challenges. To illustrate this point, in Fig. 2 we show the closest and farthest Robot test set segmentations for the die and pan.

Fig. 2. ODMS Robot Test Set Segmentation Examples. The small die segmentation (top) has fragments of other objects in the closest view (left) and completely misses the die in the farthest view (right). On the other hand, the larger pan segmentation (bottom) misses parts of the handle that are out of the image in the closest view (left) but is fairly accurate in the farthest view (right)

References

1. Griffin, B., Florence, V., Corso, J.J.: Video object segmentation-based visual servo control and object depth estimation on a mobile robot. In: IEEE Winter Conference on Applications of Computer Vision (WACV) (2020)