Supplemental Material:
Augmenting Depth Estimation with Geospatial Context

Scott Workman
DZYNE Technologies

Hunter Blanton
University of Kentucky

1. HoliCity-Overhead Dataset

We introduced the HoliCity-Overhead dataset, an extension of the the HoliCity [1] dataset that includes overhead imagery and height data. Figure 1 visualizes the coverage of the original dataset in downtown London, UK. Figure 2 shows example overhead images and aligned height map pairs (each image of size 512×512) at the different zoom levels we collected (i.e., varying ground sample distance or ground resolution).

2. Extended Results

We present additional evaluation of our methods using the HoliCity-Overhead dataset.

2.1. Height Estimation

Table 1 shows results for height estimation, generated using our method (Section 4 of the main paper). For this experiment we used the zoom level 17 imagery, which equates to a ground sample distance of approximately .74 meters per pixel over London. Though height estimation is only an intermediate task of our network and not the primary objective, our approach achieves good performance in several metrics. Figure 3 shows example height maps generated using our approach alongside the ground truth, with yellow indicating larger values.

Table 1: Height estimation results on HoliCity-Overhead.

	MAE	RMSE	RMSE log
ours (height est.)	3.333	5.225	0.470

2.2. Scale Factor Analysis

Though our method requires no computation of a scaling factor at inference, we can analyze the performance of our method in terms of scale. Figure 4 shows a scatter plot (per image) where the x-axis is median depth from ground-truth and the y-axis is median depth from prediction. Note that in median scaling, the scaling factor is estimated as the

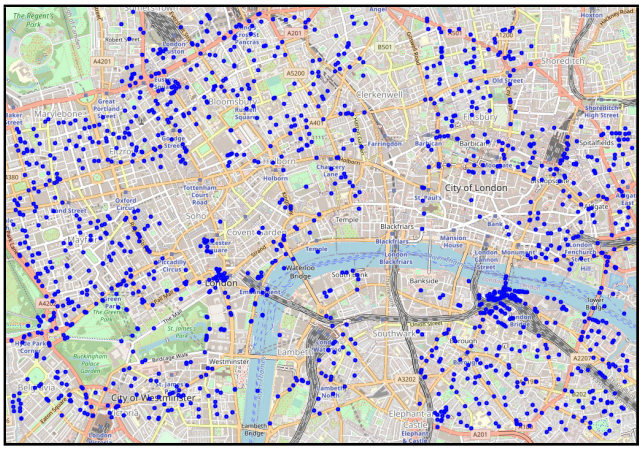

Figure 1: Visualizing the coverage of the HoliCity [1] dataset. The locations of the ground-level panoramas are shown as blue dots (subsampled).

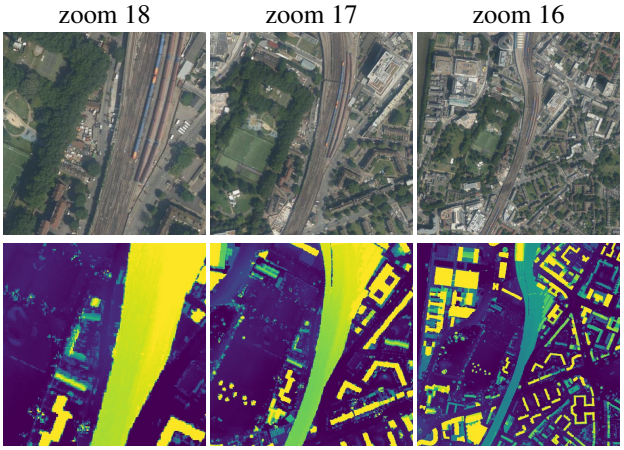

Figure 2: Example data from the HoliCity-Overhead dataset at different zoom levels (i.e., varying ground sample distance). (top) Overhead image and (bottom) height map from composite digital surface model.

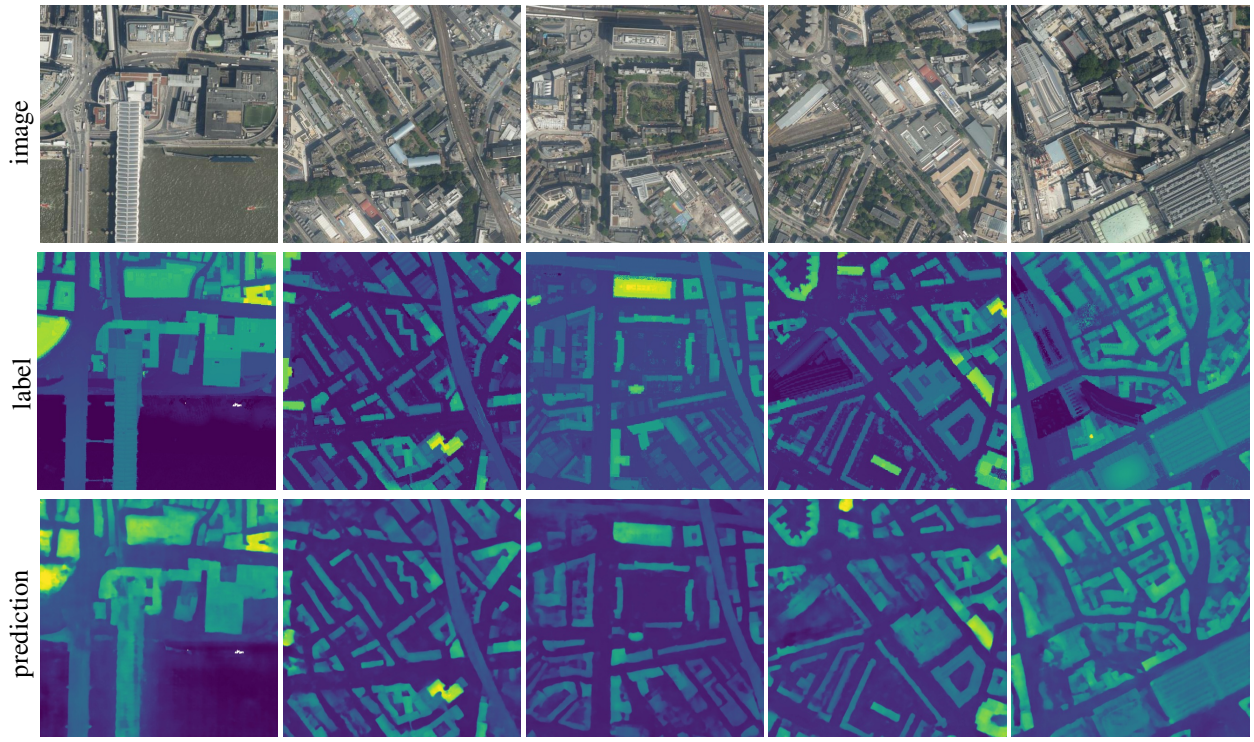

Figure 3: Example height maps generated by our approach where yellow represents larger values.

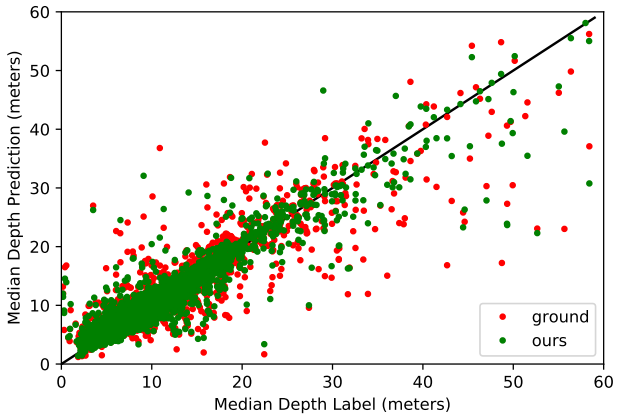

Figure 4: Visualizing median depth of ground-truth versus median depth of prediction (the ratio of which is the scale factor for median scaling) for our method versus a baseline. Each dot corresponds to an image in the HoliCity-Overhead evaluation set. Our method is generally closer to the diagonal, indicating better performance.

ratio of these two quantities. For this visualization, points closer to the diagonal indicate better alignment in scale. As observed, our approach is generally closer to the diagonal when compared against the ground-only baseline.

2.3. Impact of Ground Sample Distance

As shown in Section 5.2 of the main paper, starting from an overhead height map with larger ground sample distance leads to better performance when integrating geospatial context. This makes sense because a larger ground sample distance, but same size image, has greater spatial coverage. Here we visualize the impact of this on the generated synthetic depth panoramas. For this experiment, we use the ground-truth height map data contained in the HoliCity-Overhead dataset. Figure 5 visualizes the results. As observed, synthetic depth panoramas generated from zoom level 16 data (approx. 1.5 meters per pixel) contain objects, such as buildings, that are farther away than in zoom level 18 (approx. .4 meters per pixel).

2.4. Qualitative Results

Finally, in Figure 6 we show results generated by our method alongside results from the ground-only baseline. As observed, our approach that integrates geospatial context is better able to capture the scale of the scene. For example, in the first row, the ground-only baseline significantly underestimates the maximum depth along the road compared to our approach.

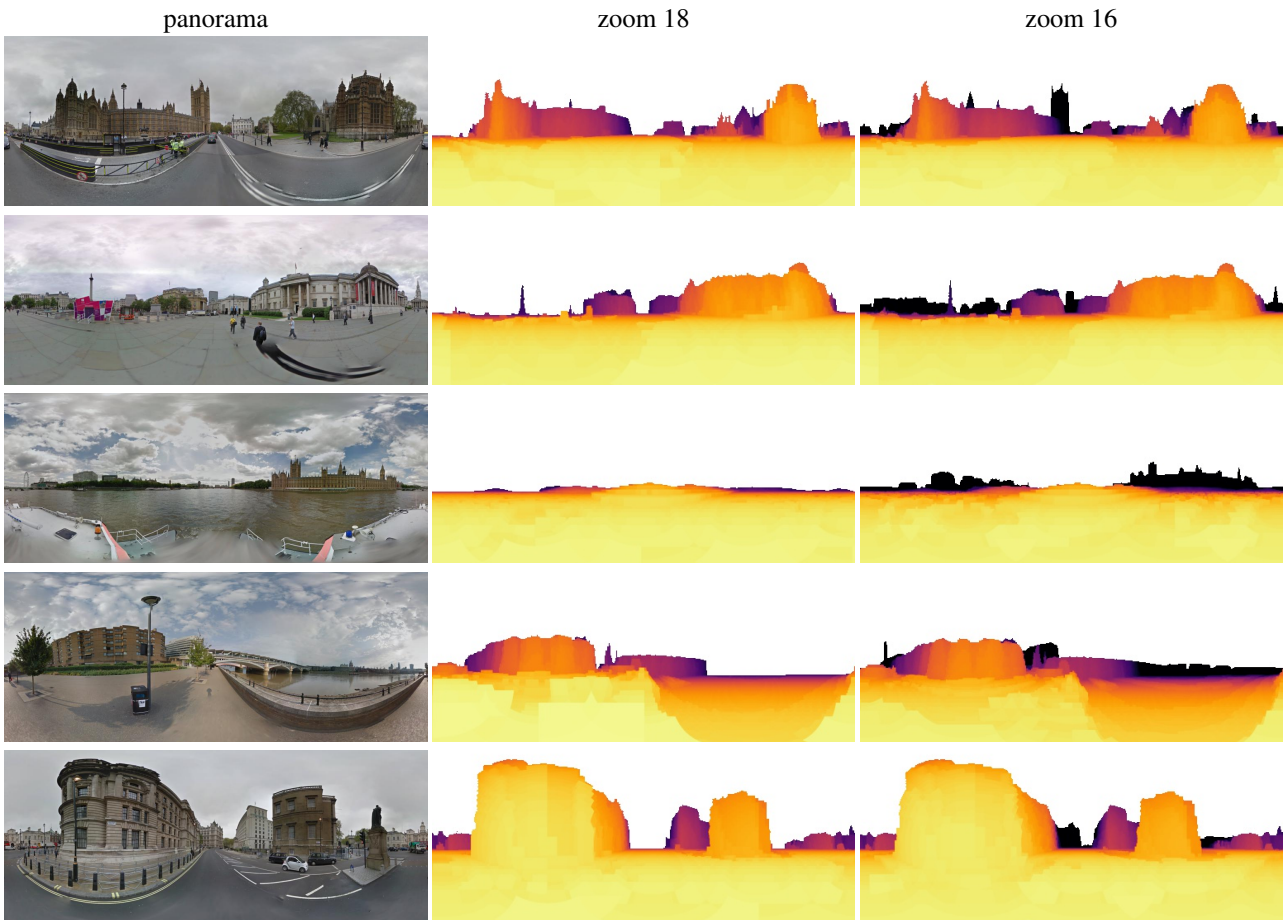

Figure 5: Visualizing the impact of ground sample distance. Starting from lower zoom levels (higher ground sample distance) increases spatial coverage, enabling capture of objects farther away in the synthetic depth panoramas. Yellow (black) values indicate smaller (larger) depths.

3. Application: Estimating Geo-Orientation

We show that our intermediate representation of scale (in the form of a synthetic depth panorama) can be used for a variety of applications, including orientation estimation. For this experiment, we use the overhead height map at zoom level 16. Given a query ground-level depth map with known geolocation but unknown orientation, we first generate the synthetic depth panorama from a co-located height map using our approach. We then perform a grid search over yaw/pitch parameters at one degree intervals, extracting the corresponding perspective depth cutout, and comparing to the query depth image. Each orientation is assigned a score using mean absolute error between the two depth maps, selecting the lowest error configuration as our prediction. Example registration results are shown in Figure 7. The ground-truth perspective cutout boundary is shown in blue and the result of our registration technique is shown in red. Despite this simple approach, the estimated orientations are quite accurate.

References

[1] Yichao Zhou, Jingwei Huang, Xili Dai, Linjie Luo, Zhili Chen, and Yi Ma. HoliCity: A city-scale data platform for learning holistic 3D structures. *arXiv preprint arXiv:2008.03286*, 2020. 1

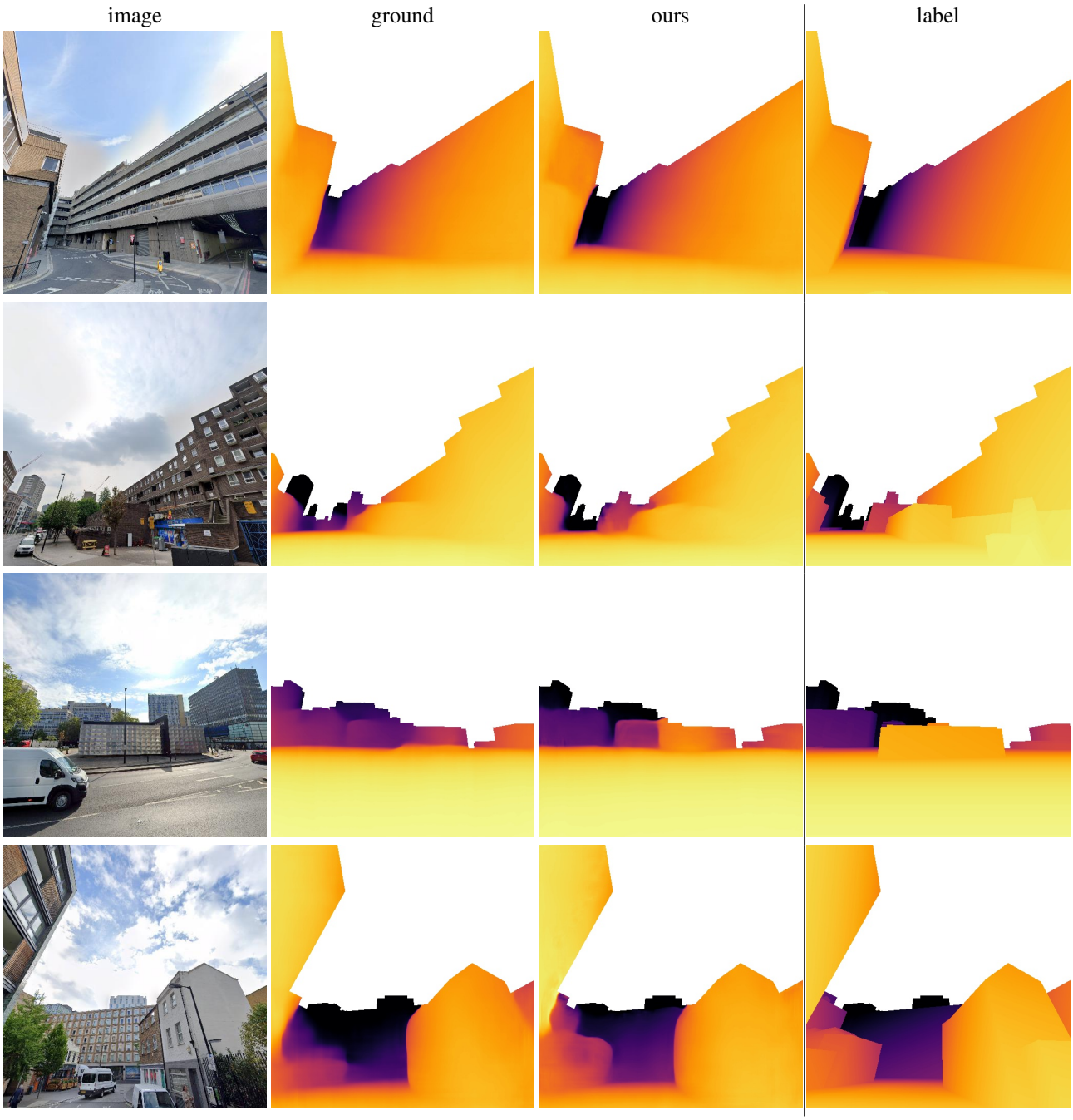

Figure 6: Qualitative results of our method compared to the ground-only baseline. Our approach, which integrates geospatial context, better captures the scale of the scene. Yellow (black) values indicate smaller (larger) depths.

Figure 7: Example results from orientation estimation. The perspective image boundary corresponding to the true orientation is shown in blue and our registration result is shown in red.