Toward Planet-Wide Traffic Camera Calibration

SUPPLEMENTARY MATERIAL

Khiem Vuong Robert Tamburo Srinivasa G. Narasimhan
{kvuong,rtamburo,srinivas}@andrew.cmu.edu
Carnegie Mellon University

1. Summary and More Results

For a thorough summary of our method and more results, please see our project website.

2. Baselines Details

In our main paper, we conducted a comprehensive comparison of our proposed method with several state-of-the-art (SOTA) automatic camera calibration approaches, specifically designed for traffic cameras. These methods include OptInOpt [1], PlaneCalib [2], DeepVPCalib [5], and the approach by Revaud et al. [7].

OptInOpt [1] and PlaneCalib [2] rely on localizing 2D landmarks using precise 3D CAD models to estimate the camera's focal length and vehicle poses. Notably, these methods were initially developed with the assumption that most vehicles in the BrnoCompSpeed [9] dataset from the Czech Republic are predominantly manufactured by Skoda. This unique dataset characteristic allowed the original implementations to establish exact 2D-3D correspondences for the keypoints. However, as highlighted in our main manuscript, this assumption isn't realistic and doesn't readily generalize to other regions or countries with diverse car models. In addressing this limitation, we drew inspiration from GSNet [4]. Our approach involved classifying each detected car instance into one of four mean shape variations, enabling the establishment of correspondences. Regarding Revaud et al.'s method [7], we used the checkpoint they provided, often referred to as the "learned method" in their work [7]. This model was exclusively trained using synthetic 3D car models, thus exhibiting limited generalization capabilities.

3. Statistics

In our 3D reconstruction pipeline, we use an average of 25 panorama images or 300 perspective images for each intersection. The most computationally intensive step involves SuperGlue feature matching [8]. Here, we match each image with $k = 50$ similar images retrieved using vocabulary tree matching, which results in an average of $15,000$ image pairs. On a single NVIDIA RTX 4090 GPU, this process consumes approximately 20 minutes. It's important to highlight that this process can be parallelized, making it scalable with the number of available GPUs. When combined with the subsequent COLMAP reconstruction step, the total average runtime for the entire 3D reconstruction workflow amounts to just under 1 hour per scene. Importantly, it's worth noting that the 3D reconstruction only needs to be performed once for each intersection. Following this initial reconstruction, any traffic camera subsequently installed at the same location can be localized with minimal runtime requirements.

4. More Qualitative Results

Our camera localization method proves versatile across diverse cameras in real-world settings. As depicted in Fig. 1, we successfully achieve both accurate 3D scene reconstruction and precise camera localization across different locations spanning multiple countries and continents. Importantly, our framework's adaptability goes beyond Google Street View, making it versatile for various street-level imagery sources [3, 6]. This enhances scalability, enabling global-scale camera calibration.

References

[1] Vojtěch Bartl and Adam Herout. Optinopt: Dual optimization for automatic camera calibration by multi-target observations. In *AVSS*, 2019. 1

[2] Vojtěch Bartl, Roman Juranek, Jakub Špaňhel, and Adam Herout. Planecalib: Automatic camera calibration by multiple observations of rigid objects on plane. In *DICTA*, 2020. 1

[3] Bing. Bing Streetside. https://www.bing.com/maps/. 1

[4] Lei Ke, Shichao Li, Yanan Sun, Yu-Wing Tai, and Chi-Keung Tang. Gsnet: Joint vehicle pose and shape reconstruction with geometrical and scene-aware supervision. In *European Conference on Computer Vision*, pages 515–532. Springer, 2020. 1

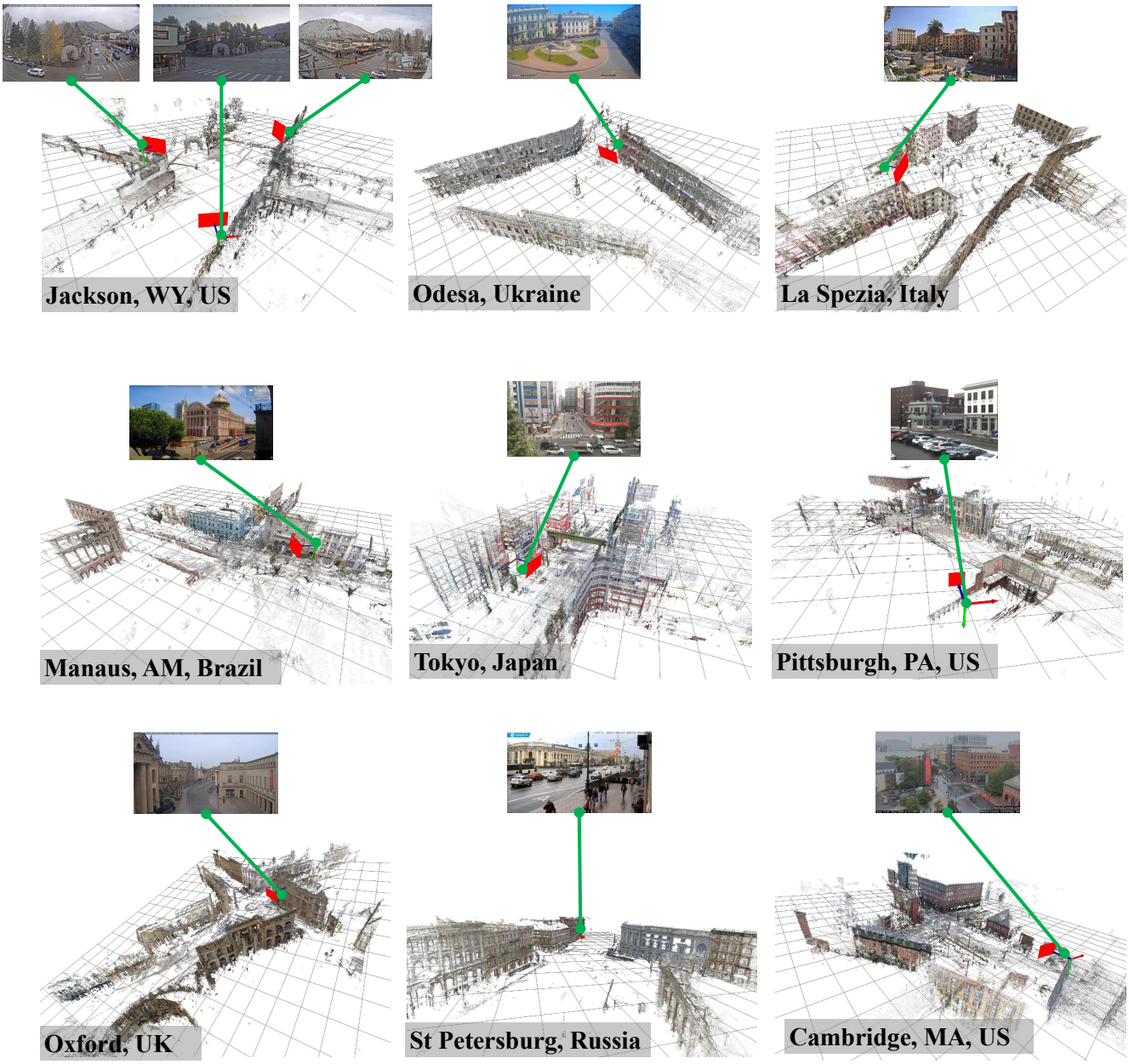

Figure 1. Additional examples demonstrate our method's robustness in reconstructing scenes and localizing cameras spanning various countries and continents (traffic camera visualized in red).

[5] Viktor Kocur and Milan Ftáčnik. Traffic camera calibration via vehicle vanishing point detection. In *ICANN*, pages 628–639. Springer, 2021. 1

[6] Mapillary. Mapillary Maps. https://www.mapillary.com/. 1

[7] Jerome Revaud and Martin Humenberger. Robust automatic monocular vehicle speed estimation for traffic surveillance. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4551–4561, 2021. 1

[8] Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superglue: Learning feature matching with graph neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4938–4947, 2020. 1

[9] Jakub Sochor, Roman Juránek, Jakub Špaňhel, Lukáš Maršík, Adam Širokỳ, Adam Herout, and Pavel Zemčík. Comprehensive data set for automatic single camera visual speed measurement. *IEEE Transactions on Intelligent Transportation Systems*, 20(5):1633–1643, 2018. 1