Homography Guided Temporal Fusion for Road Line and Marking Segmentation Appendix

Shan Wang1,2 Chuong Nguyen1 Jiawei Liu2 Kaihao Zhang2 Wenhan Luo3 Yanhao Zhang2 Sundaram Muthu1 Fahira Afzal Maken1 Hongdong Li2
1Data61, CSIRO 2Australian National University 3Sun Yat-sen University

1. Derivation of Jacobian

In this section, we present the detailed derivations for the Equation. 9, Equation. 10 and Equation. 11 of the main paper, which calculate the Jacobian of Homography Transformation w.r.t. the pitch angle $\theta$ and the roll angle $\phi$. The pitch angle $\theta$ and the roll angle $\phi$ represent the rotation of the road surface w.r.t. the plane constructed with the x-axis and z-axis of camera coordinates, as shown in Fig. 1.

$$\frac{\partial p_i}{\partial n} = \frac{\partial(K(R_i - \frac{t_n}{d})K^{-1}(p_i + 1))}{\partial n}$$

$$= \frac{\partial(KR_iK^{-1}(p_i + 1))}{\partial n} - \frac{\partial(K \frac{t_n}{d})K^{-1}(p_i + 1))}{\partial n}$$

$$= 0 - \frac{1}{d} \frac{\partial(Kt_i n^\top K^{-1}(p_i + 1))}{\partial n}$$

$$= -\frac{1}{d} Kt_i (K^{-1}(p_i + 1))^\top,$$

(1)

Derivation of Equation. 9 of the Main Paper is the Jacobian of road surface normal w.r.t. road surface normal $n$.

$$\frac{\partial n}{\partial \theta} = (\frac{\partial(- \sin \phi \cos \theta)}{\partial \theta}, \frac{\partial(- \cos \phi \cos \theta)}{\partial \theta}, \frac{\partial(\sin \theta)}{\partial \theta})$$

$$= (\sin \phi \sin \theta, \cos \phi \sin \theta, \cos \theta$$

$$= (\frac{\sin \phi \cos \theta \sin \theta}{\cos \theta}, \frac{\cos \phi \cos \theta \sin \theta}{\cos \theta}, \cos \theta)$$

$$= (-\frac{n_1 n_3}{\cos \theta} - \frac{n_2 n_3}{\cos \theta}, \cos \theta)$$

$$= (-\frac{n_1 n_3}{\sqrt{1 - n_3^2}}, -\frac{n_2 n_3}{\sqrt{1 - n_3^2}}, \sqrt{1 - n_3^2}),$$

Derivation of Eq. 11 of the Main Paper is the Jacobian of road surface normal w.r.t. roll angle $\phi$.

$$\frac{\partial n}{\partial \phi} = (\frac{\partial(- \sin \phi \cos \theta)}{\partial \phi}, \frac{\partial(- \cos \phi \cos \theta)}{\partial \phi}, \frac{\partial(\sin \theta)}{\partial \phi})$$

$$= (-\cos \phi \cos \theta, \sin \phi \cos \theta, 0)$$

$$= (n_2, -n_1, 0),$$

2. Performance with Estimated Camera Extrinsic

One limitation of our method is that it requires additional camera extrinsic information. Although it is easy to obtain in a real vehicle system, it may not be available for some datasets. To explore our performance under extrinsic unknown settings, we estimated the camera extrinsic using COLMAP [9] and used the estimated extrinsic to evaluate the proposed method. However, the ApolloScape dataset’s larger camera elevation angle caused the key points obtained by COLMAP to be further away from the vehicle, resulting in noisy camera poses. Despite these less accurate poses, our method still achieved plausible performances (“w/ COLMAP Extrinsic” in Tab. 1).

3. Performance with Estimated Homography

We also use FindHomography function from OpenCV package to directly estimate the homography transforma-
tion matrix. The estimations are less accurate than our HomoGuide, leading to inferior lane segmentation results (“w/ OpenCV Homography” of Tab. 1).

Table 1: Comparison of Estimated Extrinsic and OpenCV Homography on ApolloScape Dataset

<table>
<thead>
<tr>
<th></th>
<th>18 mIoU↑</th>
<th>36 mIoU↑</th>
</tr>
</thead>
<tbody>
<tr>
<td>HomoFusion (ours)</td>
<td>59.3</td>
<td>35.9</td>
</tr>
<tr>
<td>w/ COLMAP Extrinsic</td>
<td>56.2</td>
<td>34.8</td>
</tr>
<tr>
<td>w/ OpenCV Homography</td>
<td>47.9</td>
<td>29.4</td>
</tr>
</tbody>
</table>

4. Challenging Scenarios

In this section, we evaluate the performance of our method on challenging scenarios. As we utilize the LM algorithm for road surface normal vector estimation, there is a convex range of the algorithm that can result in failure due to excessive re-projection errors. To further investigate this scenario, we selected the top 100 inputs with the highest residuals after optimization. We compared our approach with and without RSNE and found that incorporating the RSNE led to improved performance (18 mIoU: 53.3 and 52.5, respectively), even in cases where the optimization has a high possibility of being non-convex.

Additionally, we investigate whether our method successfully estimates the road surface normal vector under uphill/downhill road scenarios. Fig. 2 demonstrates that the normal is accurately estimated in uphill scenarios, resulting in good alignment of road marks across multiple frames.

![Figure 2: Correctly fused uphill frames using estimated normal.](image)

5. Dataset Selection Criteria

Our method relies on borrowing information from adjacent frames, and therefore datasets that do not include continuous frames with a common field of view, such as CeyMo [6], CULane [10], and VPYNet [8], are unsuitable for our purposes. These datasets do not provide the necessary information for our method to work effectively, as they lack the continuity required for the information to be meaningfully borrowed across frames. In addition, although vehicle pose is free information that we can exploit, transitional lane mark detection datasets, such as LLAMAS [2], SDLane [7] and VIL-100 [14], do not provide camera intrinsic and extrinsic information, nor do they provide GPS/IMU data that could be used to calculate camera global extrinsic information. These datasets are unsuitable for our method because they lack the necessary information for our approach to work effectively. Furthermore, our method is a segmentation task, and as such, datasets that do not provide segmentation masks, such as OpenLane [3], Tusimple [10] are also unsuitable. Without segmentation masks, it is impossible to accurately determine the boundaries of lane markings in the image, making it difficult to apply our method effectively. While the Waymo Open Dataset [12] primarily targets general panoptic segmentation, our lane marker segmentation task focuses on only two categories—lane markers and road markers—limiting its utility. Our comparison solely involves CFFM [11] on this dataset, highlighting our approach’s enhanced efficacy in Table 2.

Table 2: Comparison on Waymo Open Dataset

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3 mIoU↑</td>
<td>63.22</td>
<td>66.45</td>
</tr>
</tbody>
</table>

Given these limitations, our primary experimentation is carried out on the ApolloScape dataset [5], as it offers the necessary information for our approach. In addition, we created an artificial dataset called ApolloScape Night from the ApolloScape dataset using a cross-domain generation network [1]. This dataset allows us to evaluate the effectiveness of our method under challenging lighting conditions. We conducted experiments on these datasets to demonstrate the effectiveness of our method across various domains, including daytime and nighttime driving scenarios. The results show that our approach achieves SOTA performance on these datasets, confirming its effectiveness and suitability for real-world lane detection applications.

6. Visualization on Real Night/Rain Scenes

We verify the effectiveness of our proposed method in challenging real-world environments, e.g. dark and rainy situations. Fig. 3 and Fig. 4 present the segmentation results of our proposed method on real images taken at night and in rainy conditions, respectively. Despite being trained only on the artificial ApolloScape Night dataset, which simulates road conditions at night, our proposed method successfully segments various road lines and markings in real images.

7. Visualization on HomoGuide

This section complements Section. 4.6 of the Main Paper on the “Impact of HomoFusion”. Fig. 7 compares our proposed method with and without Homo Guide. It illustrates
Figure 3: Segmentation results (Right) of real night frames (Left) using the model trained on the artificial ApolloScape Night dataset. Our proposed method accurately segments various road lines and markings, such as dot lines, solid lines, crosswalks, and straight arrows.

Figure 4: Segmentation results (Right) of real rainy day frames (Left) using the model trained on the artificial ApolloScape Night dataset. The proposed method accurately segments various road lines and markings, including dot lines, solid lines, double yellow lines, stop lines, and crosswalks.

that incorporating HomoGuide enables our method to accurately classify road lines and markings even under adverse conditions such as occlusion, road reflection, and poor light conditions.

8. Visualization of Water Hazard Detection

This section complements Section. 4.6 of the Main Paper on the “Application to Another Task”, we provide a visualization of the predicted segmentation and the input frames transformed by Homography in Fig. 5 and Fig. 6, respectively. These figures demonstrate that our proposed HomoFusion can align ground surfaces well enough to improve the performance of detection of flat objects on the ground.

Figure 5: Current frame, prediction and ground truth masks of water puddle segmentation. The 1st and 2nd rows are from the Off-road dataset, and the 3rd and 4th rows are from the On-road dataset.

Figure 6: Input frames transformed to the current frame by homography using the estimated road surface normal vectors and camera movement. The fused image of all eight frames is shown in the bottom right corner.

9. Extra Qualitative Examples

Fig. 8 demonstrates that our proposed method can accurately segment the road lines and markings of various categories.
Furthermore, we provide additional qualitative examples of our proposed method and state-of-the-art (SOTA) algorithms, including (a) IntRA-KD [4], (b) SegFormer [13], and (c) CFFM [11] and (d) MMA-Net [14], on ApolloScape [5] in Fig. 9 and ApolloScape Night in Fig. 10. Our method demonstrates better true-positive predictions and fewer false-positive predictions in both scenarios. Additionally, our method accurately predicts more precise boundaries even in challenging glare and poor lighting conditions.

References


[3] Li Chen, Chonghao Sima, Yang Li, Zehan Zheng, Jiajie Xu, Xiangwei Geng, Hongyang Li, Conghui He, Jianping Shi,
Figure 9: Extra qualitative comparison of our proposed method with SOTA methods on the ApolloScape [5] dataset. Yellow boxes highlight the differences and enlarge the target area for better visibility. Red boxes indicate false-positive segmentation predictions.
Figure 10: Extra qualitative comparison of our proposed method with SOTA methods on the ApolloScape Night dataset. Yellow boxes highlight the differences and enlarge the target area for better visibility. Red boxes indicate false-positive segmentation predictions.


