Generating 3D Depth Maps for Autonomous Vehicles Using Multi-View Image Stitching

Ramtin Moslemi Amirhossein Haji Mohammad Rezaie Ahmadreza Khanari

October 20, 2025

Under the supervision of Dr. Shohreh Kasaei
Department of Computer Engineering, Sharif University of Technology

Abstract

This paper presents a method for creating three-dimensional depth maps of a vehicle's surroundings to enhance perception and navigation in autonomous vehicles. We utilize images captured from multiple vehicle-mounted cameras in different directions, stitching them to form a panoramic view. Subsequently, we generate a comprehensive depth map from this panoramic image to improve environmental understanding for autonomous driving applications. Our approach addresses challenges such as blind spots and limited fields of view, demonstrating improved scene reconstruction through iterative homography estimation and depth estimation techniques. Experiments on datasets like nuScenes and Oxford RobotCar validate the efficacy of our method.

keywords: Autonomous Vehicles, Image Stitching, Depth Estimation, Computer Vision, Panoramic Imaging

1 Introduction

Autonomous vehicles represent a pivotal advancement in modern technology, facing numerous challenges in perceiving their surrounding environment, primarily within the domain of computer vision. The algorithms employed must be highly sophisticated to minimize accidents and associated risks. Beyond autonomous systems, human drivers can also benefit from computer vision technologies, such as sensors and cameras that are now commonplace in vehicles.

A critical issue is the presence of blind spots in camera views, particularly in larger vehicles. For instance, objects like motorcycles in these blind spots may go undetected, leading to potential hazards.

To mitigate this, our work focuses on fusing images from rear, left, and right cameras using image stitching to create a panoramic view that eliminates blind spots. We then derive a depth map from this panorama to aid in collision avoidance and environmental perception for autonomous systems.

In summary, our contributions include:

1. Generating panoramic images from multi-directional vehicle camera feeds.
2. Computing depth maps from the stitched panoramas to enhance 3D scene understanding.

This system replaces traditional mirrors with cameras and sensors, paving the way for advanced autonomous driving. Experts predict widespread adoption of autonomous vehicles by the 2030s, necessitating extensive driving data collection for further refinements.

2 Related Work

We review prior works on image stitching and depth estimation relevant to autonomous vehicles.

2.1 Image Stitching

Image stitching typically involves detecting keypoints in overlapping images, matching them to compute a homography matrix, and applying projective transformations to align images [2]. To achieve natural views, panoramic images can be projected onto cylindrical or spherical surfaces for quality optimization [3].

In autonomous vehicle contexts, stitching is often applied to fisheye camera images. For example, [4] employs intensity-based sum-of-squares minimization to refine homography for dual-fisheye lenses.

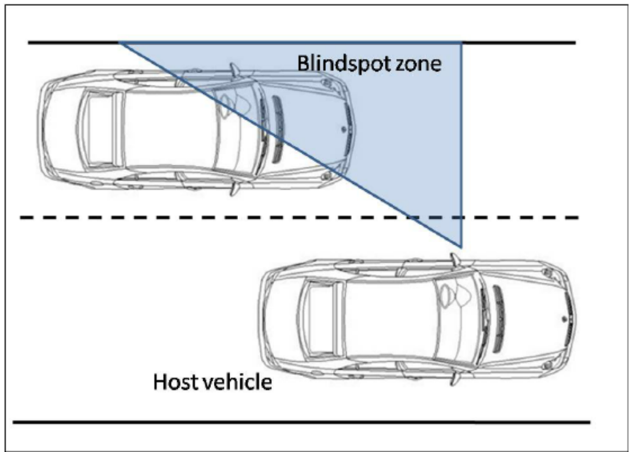

Figure 1: Illustration of a blind spot zone where objects may be undetectable [1].

2.2 Depth Estimation

Depth estimation methods largely rely on deep learning for monocular images. The MonoDepth2 model [5] uses self-supervised learning with refined loss terms and masks for accurate depth in vehicular scenes [6].

Other approaches incorporate fisheye images with sparse LiDAR supervision [7, 8]. The PanoDepth model [9] processes 360-degree panoramas, generating undistorted views for depth estimation using deep networks.

3 Proposed Method

3.1 Image Stitching

We compute homography matrices to stitch images but face challenges: unknown camera intrinsics/extrinsics from datasets and limited fields of view requiring multiple cameras for 360-degree coverage.

To address homography estimation, we use SIFT for keypoint matching in non-ideal conditions. Datasets constrain our options; ideally, custom setups would ensure overlaps, but we leverage existing ones like nuScenes.

For robustness, we iteratively estimate homography across consecutive frames, updating via exponentially weighted moving average:

$$H = \alpha H' + (1 - \alpha)H, \quad \alpha = \frac{1}{\text{cnt}} \tag{1}$$

where *cnt* is the frame counter. If drift occurs, a sliding window refines updates. Erroneous matrices (norm difference > 100) are discarded, using prior averages for stitching.

3.2 Depth Estimation

We categorize depth methods into stereo vision and deep learning, favoring the latter for monocular datasets. Using nuScenes models, we estimate object depths. For implementation, slight camera rotations align with provided data. Object detection integrates depth for 3D bounding boxes.

4 Datasets

4.1 nuScenes

nuScenes [11] comprises 1.4M images, 390K LiDAR sweeps, and annotations from 1000 scenes in Boston and Singapore, selected for diversity in urban driving challenges.

4.2 Oxford RobotCar

Collected over 1000 km in Oxford (2014-2015), it includes 20M images from 6 cameras, LiDAR, GPS, under varied weather and conditions [10].

4.3 CARLA Simulator

CARLA [12] enables customizable sensor simulations for dataset generation. We use it to create tailored multi-view images [13].

5 Experiments and Results

Refer to accompanying notebooks for full implementations. Sample stitching results for the vehicle's left side are shown below.

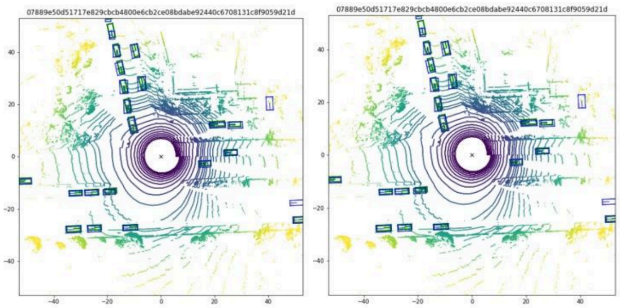

Figure 2: Depth estimation example: 'x' denotes vehicle position, blue rectangles indicate detected cars.

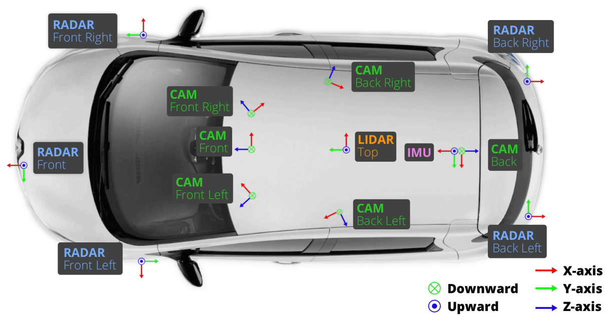

Figure 3: Camera setup in nuScenes [11].

6 Challenges

Dataset volumes exceed 15GB, limiting usage to subsets. Limited FoV reduces keypoint overlaps, degrading homography accuracy.

7 Conclusion and Future Work

Our method advances autonomous vehicle perception by integrating stitching and depth mapping. Future efforts could focus on commercial integration, collecting extensive data for enhanced systems.

References

[1] Blind spot zone. [Online]. Available: https://www.researchgate.net/figure/The-blindspot-zone-description-We-define-the-blindspot-of-a-driver-as-the-zone-he-can_fig1_221355854

[2] M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," Int. J. Comput. Vis., vol. 74, no. 1, pp. 59–73, 2007.

[3] C.-C. Lin et al., "Adaptive as-natural-as-possible image stitching," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2015, pp. 1155–1163.

[4] T. Ho et al., "360 degree video stitching for dual fisheye cameras," in Proc. IEEE Int. Conf. Image Process. (ICIP), 2017, pp. 2174–2178.

[5] C. Godard et al., "Digging into self-supervised monocular depth estimation," in Proc. IEEE/CVF Int. Conf. Comput. Vis., 2019, pp. 3828–3838.

[6] A. Johnston and G. Carneiro, "Self-supervised monocular trained depth estimation using self-attention and discrete disparity volume," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2020, pp. 4756–4765.

[7] V. R. Kumar et al., "Monocular fisheye camera depth estimation using sparse lidar supervision," in Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC), 2018, pp. 2853–2858.

[8] V. R. Kumar et al., "FisheyeDistanceNet: Self-supervised scale-aware distance estimation using monocular fisheye camera for autonomous driving," in Proc. IEEE Int. Conf. Robot. Autom. (ICRA), 2020, pp. 5741–5748.

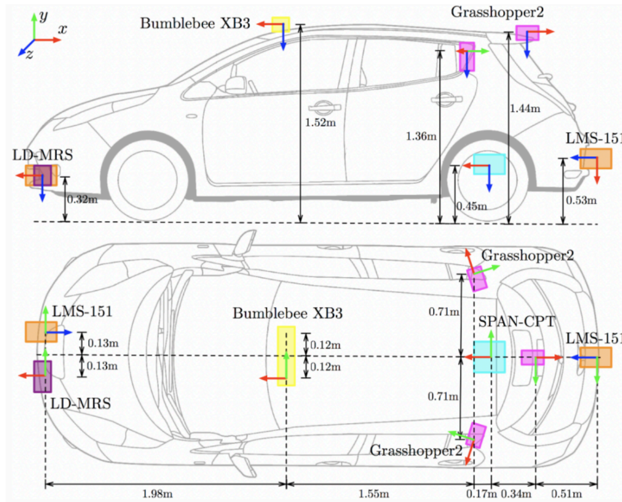

Figure 4: Camera placements in Oxford RobotCar [10].

[9] Y. Li et al., "PanoDepth: A two-stage approach for monocular omnidirectional depth estimation," in Proc. Int. Conf. 3D Vis. (3DV), 2021, pp. 71–80.

[10] W. Maddern et al., "1 year, 1000 km: The Oxford RobotCar dataset," Int. J. Robot. Res., vol. 36, no. 1, pp. 3–15, 2017.

[11] H. Caesar et al., "nuScenes: A multimodal dataset for autonomous driving," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2020, pp. 11621–11631.

[12] A. Dosovitskiy et al., "CARLA: An open urban driving simulator," in Proc. Conf. Robot Learn., 2017, pp. 1–16.

[13] R. Nett. (2020). CARLASim. [Online]. Available: https://github.com/rnett/CARLASim

Figure 5: Samples from CARLA [13].

Figure 6: Image stitching results for left-side views.