




Freespace Optical Flow Modeling for Automated Driving

Yi Feng , Graduate Student Member, IEEE, Ruge Zhang , Jiayuan Du ,
Qijun Chen , Senior Member, IEEE, and Rui Fan , Member, IEEE

Abstract—Optical flow and disparity are two informative visual features for autonomous driving perception. They have been used for a variety of applications, such as obstacle and lane detection. The concept of “U-V-Disparity” has been widely explored in the literature, while its counterpart in optical flow has received relatively little attention. Traditional motion analysis algorithms estimate optical flow by matching correspondences between two successive video frames, which limits the full utilization of environmental information and geometric constraints. Therefore, we propose a novel strategy to model optical flow in the collision-freespace (also referred to as drivable area or simply freespace) for intelligent vehicles, with the full utilization of geometry information in a 3-D driving environment. We provide explicit representations of optical flow and deduce the quadratic relationship between the optical flow component and the vertical coordinate. Through extensive experiments on several public datasets, we demonstrate the high accuracy and robustness of our model. In addition, our proposed freespace optical flow model boasts a diverse array of applications within the realm of automated driving, providing a geometric constraint in freespace detection, vehicle localization, and more. We have made our source code publicly available at <https://mias.group/FEOF>.

Index Terms—Automated driving, autonomous driving perception, freespace, optical flow, vehicle localization.

Manuscript received 17 February 2023; revised 15 July 2023; accepted 20 July 2023. Recommended by Technical Editor X. Zhang and Senior Editor X. Chen. This work was supported in part by the National Key R&D Program of China under Grant 2020AAA0108100, in part by the National Natural Science Foundation of China under Grant 62233013, in part by the Science and Technology Commission of Shanghai Municipal under Grant 22511104500, and in part by the Fundamental Research Funds for the Central Universities. (Yi Feng and Ruge Zhang are co-first authors.) (Corresponding author: Rui Fan.)

Yi Feng, Jiayuan Du, Qijun Chen, and Rui Fan are with the College of Electronics and Information Engineering, Shanghai Research Institute for Intelligent Autonomous Systems, State Key Laboratory of Intelligent Autonomous Systems, and Frontiers Science Center for Intelligent Autonomous Systems, Tongji University, Shanghai 201804, China (e-mail: fengyi@ieee.org; dujiayuan@tongji.edu.cn; qjchen@tongji.edu.cn; rui.fan@ieee.org).

Ruge Zhang is with the High-Performance Computer Research Center, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China (e-mail: zhangruge23@mailsucas.ac.cn).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TMECH.2023.3300729>.

Digital Object Identifier 10.1109/TMECH.2023.3300729

I. INTRODUCTION

WITH the advancements in sensor technology and navigation positioning systems, the advanced driver assistance system (ADAS) has emerged as a crucial component of intelligent vehicles [1]. It utilizes multiple sensors to monitor the surrounding environment and provides real-time information to the front end of the automated driving system [2]. Simultaneously, the emergence and evolution of various computer vision algorithms have driven continuous improvements in the accuracy and performance of the environmental perception functionality in ADAS systems [3]. Motion analysis poses a fundamental and challenging problem in environmental perception [4], which entails the estimation of 2-D or 3-D object motion using dynamic scene sequences captured across multiple successive video frames. Optical flow estimation focuses specifically on analyzing 2-D motion, whereas scene flow estimation deals with 3-D motion analysis [5], [6]. Optical flow and scene flow estimations hold profound significance in establishing high-level cognitive ability in 3-D scene understanding and play a pivotal role in downstream ADAS applications, such as mobile robot navigation [7] and semantic scene parsing [3].

Depth estimation and object tracking are two crucial tasks in the ADAS. Depth estimation typically involves using a stereo rig to acquire disparity information [8], which is inversely proportional to the depth. In autonomous driving scenarios, the depth map often exhibits gradual changes along the V -axis of the image plane due to the geometric features of the road surface [9]. On the other hand, object tracking is commonly accomplished or aided by optical flow information, which describes the pixel-level relationship by a 2-D motion field based on luminosity consistency. Fig. 1 illustrates the difference and relationship between optical flow and disparity in autonomous driving scenarios. I_t^l and I_t^r represent the RGB images obtained by the left and right cameras at time t , respectively, while I_{t+1}^l represents the image captured by the left camera at time $t + 1$. The optical flow map $F \in \mathbb{R}^{H \times W \times 2}$ is obtained by calculating the displacements of corresponding pixels between two successive video frames, where F_u and F_v store the horizontal and vertical displacements, respectively. This process introduces a 3-D translation and a slight rotation from the perspective of the camera. In contrast, the disparity map represents corresponding pixels that are horizontally shifted between the stereo image pair. In this process, it is considered that a monocular camera produces a horizontal translation b (the stereo rig baseline)

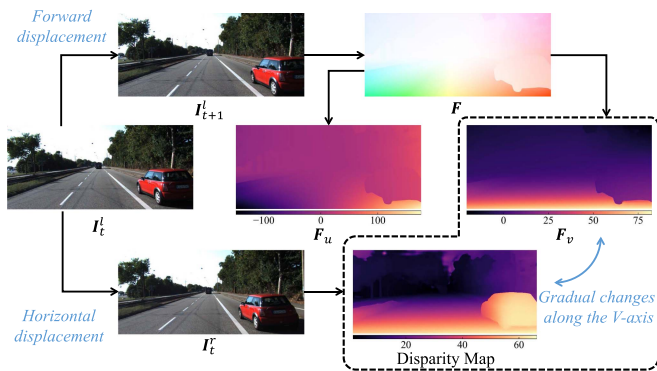


Fig. 1. Difference and relationship between optical flow and disparity.

within a very short period of time. As shown in Fig. 1, both the disparity map and F_v exhibit similar distributions: They both change gradually along the V -axis.

Hu et al. [10] conducted a comprehensive analysis of the relationship between disparity and pixel coordinates, which can be expressed as follows:

$$\begin{cases} u = \Phi(x^w, \Delta) + u_0 \\ v = \Psi(z^w, \Delta) + v_0 \end{cases} \quad (1)$$

where $\mathbf{p}^w = [x^w, y^w, z^w]^\top$ represents a 3-D point in the world coordinate system, $\mathbf{p} = [u, v]^\top$ denotes its projection in the left image, $\mathbf{p}_0 = [u_0, v_0]^\top$ represents the coordinates of the principal point, Φ and Ψ are two linear functions, and Δ denotes the disparity of \mathbf{p} . In addition, the geometric characteristics in 3-D space can be abstracted into plane equations with various expressions. By combining these plane equations with (1), disparity maps can be mapped to the U -Disparity or V -Disparity domains using cumulative voting algorithms [11].

This article is motivated by two observations. First, F_v exhibits a similar distribution to that of the disparity map. Second, the physical meanings of optical flow and disparity are highly similar, as both describe the coordinate offset of corresponding pixels between two successive video frames. Inspired by the U - V -Disparity work mentioned above, we aim to explore the regularities in the distribution of freespace optical flow through the mathematical modeling of the 3-D driving environment and the integration of camera parameters and odometry information.

The rest of this article is organized as follows. Section II discusses two types of optical flow definitions and relevant literature. In Section III, we derive the mathematical relationship between an arbitrary on-road pixel \mathbf{p} and its optical flow $\mathbf{F}(\mathbf{p})$ for two different types of optical flow data formats. In Section IV, we present the experimental results concerning model validation and robustness evaluation. Section V provides a detailed discussion of various applications based on our proposed model. Finally, Section VI concludes this article.

II. LITERATURE REVIEW

Optical flow can be defined in two different ways. Traditional optical flow algorithms consider optical flow as the partial derivative of image intensity with respect to time [12]. This

definition characterizes optical flow as an apparent velocity distribution of brightness patterns in the current image [13], [14]. However, this definition assumes small interframe displacements and short time intervals, which may not always hold true in real-world scenarios. On the other hand, data-driven optical flow estimation methods rely on large and accurate optical flow datasets to enhance the performance of neural networks [15], [16], [17], [18]. These datasets define optical flow ground truth as the displacement vector of corresponding pixels between two adjacent frames. This is because that the instantaneous velocity of pixels cannot be directly measured, and calibration errors are inevitable [19].

In accordance with the two aforementioned definitions, optical flow approaches can be categorized as prior knowledge-driven or data-driven. The former ones, exemplified by Horn and Schunck's flow estimation framework [13], aim to minimize distortions in optical flow and prefer solutions that exhibit smoothness. Nevertheless, these approaches assume smoothness in optical flow across the entire image and the invariance of pixel intensity between frames, making them difficult to apply in real-world scenarios. Following Horn's work, Lucas and Kanade [14] introduced a local invariance constraint to estimate sparse optical flow. However, when the input data are noisy, this method performs poorly due to its reliance on the least squares criterion. To address illumination variation problems, Brox et al. [20] introduced a gradient constancy assumption. However, this assumption is only effective for linear illumination changes and is not capable of handling large vehicle displacements.

Data-driven algorithms, such as FlowNet [21], employ deep neural networks and large optical flow datasets to enhance the accuracy of optical flow estimation. These methods yield superior results compared to traditional approaches. However, they may not fully exploit the environment geometry information, relying heavily on hardware capabilities and the quality of optical flow datasets.

This article builds upon the two definitions of optical flow and aims to systematically construct freespace optical flow models with geometry constraints. By deriving the explicit relationship between an on-road pixel \mathbf{p} and its optical flow $\mathbf{F}(\mathbf{p})$, we effectively address the problems mentioned above.

III. METHODOLOGY

An example of the monocular system mounted on autonomous vehicles for environmental perception is illustrated in Fig. 2. We assume that the monocular system fixed above the vehicle is a basic pinhole camera model with the camera intrinsic matrix \mathbf{K} , and the vehicle with the monocular camera can be considered as a rigid body.

In our model, three coordinates are considered: Vehicle coordinate system (VCS), camera coordinate system (CCS), and pixel coordinate system (PCS). Due to the rigid connection between the monocular camera and the vehicle body, we place the CCS origin at the same position as the VCS origin, with the Y -axis pointing vertically downward and the Z -axis parallel to the vehicle's motion direction. Taking into account the potential

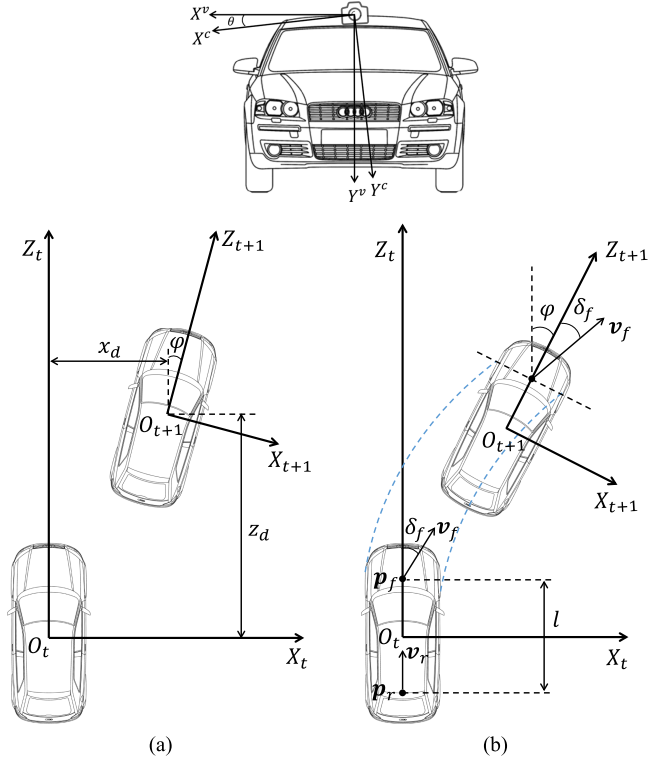


Fig. 2. Illustration of the two models. (a) Displacement-based model. (b) Velocity-based model.

mounting error and centrifugal force experienced when a car turns, there exists a roll angle θ along the Z -axis between the VCS and the CCS. Therefore, a 3-D point $\mathbf{p}_t^c = [x_t^c, y_t^c, z_t^c]^\top$ in the CCS at time t can be linked to a 3-D point $\mathbf{p}_t^v = [x_t^v, y_t^v, z_t^v]^\top$ in the VCS using $\mathbf{p}_t^c = \mathbf{R}_\theta \mathbf{p}_t^v$, where

$$\mathbf{R}_\theta = \begin{bmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

is the rotation matrix. Moreover, we consider the road surface as a horizontal plane, because the unevenness of the road and jolts while driving are negligible compared to the scale of the camera's field of view. Therefore, the road surface can be described by a plane equation: $y^v = h$, where h is the mounting height of the camera from the ground.

Our model establishes the relationship between optical flow and pixel coordinates, explicitly representing the optical flow with respect to vehicle poses and camera parameters. As discussed in Section II, there are two distinct definitions of optical flow. Therefore, we derive two different forms of the freespace optical flow model: the displacement-based freespace optical flow model, represented by \mathbf{f}_d , and the velocity-based freespace optical flow model, represented by \mathbf{f}_v . The choice between the two models should be made based on the optical flow data format. The displacement-based model is applicable when the optical flow is encoded in interframe displacement, whereas the velocity-based model is applicable when the optical flow is encoded in pixel apparent velocity.

A. Displacement-Based Freespace Optical Flow Modeling

As mentioned in Section II, optical flow is commonly defined as the displacement between two adjacent frames in most public datasets, such as Sintel [22], Flying Chairs [21], KITTI [19], [23], and Middlebury [24]. Therefore, we first establish a displacement-based freespace optical flow model, which explicitly describes the relationship between a freespace point \mathbf{p} in the PCS and its optical flow \mathbf{f}_d .

For an arbitrary 3-D point $\mathbf{p}_t^v = [x_t^v, y_t^v, z_t^v]^\top$ in the VCS, it can be projected to $\mathbf{p}_t = [u_t, v_t]^\top$ in the PCS of the video frame captured at time t as follows:

$$z_t^c \tilde{\mathbf{p}}_t = \mathbf{K} \mathbf{R}_\theta \mathbf{p}_t^v \quad (3)$$

where $z_t^c = z_t^v$ is the depth of point \mathbf{p} in the CCS of the frame at time t , and $\tilde{\mathbf{p}}_t$ represents the homogeneous coordinates of \mathbf{p}_t . Combining road surface equation ($y^v = h$) with (3), we can calculate \mathbf{p}_t^v with respect to \mathbf{p}_t :

$$\mathbf{p}_t^v = h \begin{bmatrix} \frac{f_y(u_t - u_0) \cos \theta - f_x(v_t - v_0) \sin \theta}{f_y(u_t - u_0) \sin \theta + f_x(v_t - v_0) \cos \theta} \\ 1 \\ \frac{f_x f_y}{f_y(u_t - u_0) \sin \theta + f_x(v_t - v_0) \cos \theta} \end{bmatrix} \quad (4)$$

where f_x and f_y are the horizontal and vertical camera focal lengths, respectively. In order to simplify (4), we define λ_1 and λ_2 as follows:

$$\begin{cases} \lambda_1 = \frac{u_t - u_0}{f_x} \sin \theta + \frac{v_t - v_0}{f_y} \cos \theta \\ \lambda_2 = \frac{u_t - u_0}{f_x} \cos \theta - \frac{v_t - v_0}{f_y} \sin \theta \end{cases} \quad (5)$$

Therefore, (4) can be rewritten as follows:

$$\mathbf{p}_t^v = \frac{h}{\lambda_1} \begin{bmatrix} \lambda_2 \\ \lambda_1 \\ 1 \end{bmatrix} \quad (6)$$

Since the camera mounted on the vehicle typically has a fixed distance to the ground plane, only the yaw angle φ (rotation around the Y -axis) and a translation $\mathbf{d} = [x_d, 0, z_d]^\top$ in the X -direction and Z -direction, respectively, should be considered. Therefore, we can obtain the following relationship:

$$\mathbf{p}_{t+1}^c = \mathbf{R}_\theta \mathbf{R}_\varphi (\mathbf{p}_t^v - \mathbf{d}) \quad (7)$$

where

$$\mathbf{R}_\varphi = \begin{bmatrix} \cos \varphi & 0 & -\sin \varphi \\ 0 & 1 & 0 \\ \sin \varphi & 0 & \cos \varphi \end{bmatrix} \quad (8)$$

represents the rotation with respect to φ . Therefore, the perspective projection for the second frame is as follows:

$$z_{t+1}^c \tilde{\mathbf{p}}_{t+1} = \mathbf{K} \mathbf{p}_{t+1}^c = \mathbf{K} \mathbf{R}_\theta \mathbf{R}_\varphi (\mathbf{p}_t^v - \mathbf{d}), \quad (9)$$

where $z_{t+1}^c = (x_t^v - x_d) \sin \varphi + (z_t^v - z_d) \cos \varphi$ is the depth of point \mathbf{p}_{t+1}^c . Combining (3) and (9) yields

$$\begin{bmatrix} \mathbf{f}_d \\ 0 \end{bmatrix} = \mathbf{K} \mathbf{R}_\theta \left(\frac{1}{z_{t+1}} \mathbf{R}_\varphi (\mathbf{p}_t^v - \mathbf{d}) - \frac{1}{z_t} \mathbf{p}_t^v \right). \quad (10)$$

By plugging (6) into (10), we deduce the displacement-based freespace optical flow \mathbf{f}_d as follows:

$$\mathbf{f}_d = \mathbf{M} \begin{bmatrix} \frac{\lambda_3 \cos \varphi - \lambda_4 \sin \varphi}{\lambda_3 \sin \varphi + \lambda_4 \cos \varphi} - \lambda_2 \\ \frac{\lambda_1 h}{\lambda_3 \sin \varphi + \lambda_4 \cos \varphi} - \lambda_1 \end{bmatrix} \quad (11)$$

where

$$\mathbf{M} = \begin{bmatrix} f_x \cos \theta & f_x \sin \theta \\ -f_y \sin \theta & f_y \cos \theta \end{bmatrix}$$

$$\lambda_3 = \lambda_2 h - \lambda_1 x_d$$

$$\lambda_4 = h - \lambda_1 z_d. \quad (12)$$

Therefore, we derive the explicit relationship between the optical flow and the pixel coordinates incorporating odometry and camera parameters.

B. Velocity-Based Optical Flow Modeling

As mentioned in Section II, the optical flow has also been defined as the (apparent or instantaneous) velocities of movement of brightness patterns in an image. Therefore, we also develop a velocity-based freespace optical flow model to accommodate such data format.

To derive the velocity-based freespace optical flow model, we introduce the Ackermann steering geometry [25] for four-wheel autonomous vehicle kinematics modeling. As illustrated in Fig. 2(b), δ_f denotes the steering angle of the front wheel, $\mathbf{p}_r = [x_r, z_r]^\top$ and $\mathbf{p}_f = [x_f, z_f]^\top$ respectively represent the coordinates of vehicle's rear and front axle centers, $\mathbf{v}_r = [\dot{x}_r, \dot{z}_r]^\top$ and $\mathbf{v}_f = [\dot{x}_f, \dot{z}_f]^\top$ respectively denote the velocities of the rear and front axles ($\|\mathbf{v}_r\|_2 = \|\mathbf{v}_f\|_2 = v_r$), l denotes the distance between the front and rear axles, and φ represents the vehicle's turning angle. With no sideslip of the front and rear axles, the kinematics constraints can be expressed as follows:

$$\begin{aligned} \dot{x}_f \cos(\varphi + \delta_f) &= \dot{z}_f \sin(\varphi + \delta_f) \\ \dot{x}_r \cos \varphi &= \dot{z}_r \sin \varphi \end{aligned} \quad (13)$$

where

$$\dot{x}_r = v_r \sin \varphi, \quad \dot{z}_r = v_r \cos \varphi. \quad (14)$$

According to the positions of the front and rear axle centers, we can obtain the following expressions:

$$x_f = x_r + l \sin \varphi, \quad z_f = z_r + l \cos \varphi. \quad (15)$$

Plugging (14) and (15) into (13) results in the Ackermann kinematic model as follows:

$$\begin{cases} \dot{\varphi} = v_r \frac{\tan \delta_f}{l} \\ \dot{x}_r = v_r \sin \varphi \\ \dot{z}_r = v_r \cos \varphi \end{cases} \quad (16)$$

To derive the velocity-based freespace optical flow model \mathbf{f}_v , we assume that the time interval Δt between two adjacent video frames approaches zero:

$$\mathbf{f}_v = \lim_{\Delta t \rightarrow 0} \frac{\mathbf{f}_d}{\Delta t} = \lim_{\Delta t \rightarrow 0} \mathbf{M} \mathbf{q} \quad (17)$$

where

$$\mathbf{q} = \begin{bmatrix} \frac{\lambda_1(\lambda_2 z_d - x_d) \cos \Delta \varphi - (\lambda_4 + \lambda_2 \lambda_3) \sin \Delta \varphi}{(\lambda_3 \sin \Delta \varphi + \lambda_4 \cos \Delta \varphi) / \Delta t} \\ \frac{h(1 - \cos \Delta \varphi) - \lambda_3 \sin \Delta \varphi + \lambda_1 z_d \cos \Delta \varphi}{(\lambda_3 \sin \Delta \varphi + \lambda_4 \cos \Delta \varphi) / \Delta t} \end{bmatrix}. \quad (18)$$

According to the limit theorem, we have:

$$\begin{aligned} \lim_{\Delta t \rightarrow 0} \frac{\Delta x_d}{\Delta t} &= \dot{x}_r = v_r \sin \varphi = 0 \\ \lim_{\Delta t \rightarrow 0} \frac{\Delta z_d}{\Delta t} &= \dot{z}_r = v_r \cos \varphi = v_r \\ \lim_{\Delta t \rightarrow 0} \frac{\sin \Delta \varphi}{\Delta t} &= \lim_{\Delta t \rightarrow 0} \cos \Delta \varphi \dot{\varphi} = v_r \frac{\tan \delta_f}{l} \\ \lim_{\Delta t \rightarrow 0} \cos \Delta \varphi &= 1. \end{aligned} \quad (19)$$

Plugging (19) into (17) results in

$$\mathbf{f}_v = \frac{v_r}{h} \mathbf{M} \begin{bmatrix} \lambda_1 \lambda_2 - h(1 + \lambda_2^2) \frac{\tan \delta_f}{l} \\ \lambda_1^2 - \lambda_1 \lambda_2 h \frac{\tan \delta_f}{l} \end{bmatrix}. \quad (20)$$

Equations (11) and (20) establish the fundamental relationships between pixel coordinates \mathbf{p} and their optical flow $\mathbf{F}(\mathbf{p})$ in a general monocular camera system. These relationships demonstrate that optical flow can be explicitly derived from camera model parameters and vehicle poses, eliminating the need for the brightness constancy assumption [13], which can be challenging to satisfy and may lead to estimation errors. With accurate vehicle poses and camera parameters, our proposed freespace optical flow model can be effectively deployed in various downstream autonomous driving applications.

C. Simplified Optical Flow Models for Special Cases

In most driving scenarios, the vehicle moves primarily along the Z -axis with minimal rotation and little bias in the X -direction. As a result, the yaw angle φ and the offset x_d between two adjacent frames can be approximated to zero. Consequently, (11) can be simplified as follows:

$$\mathbf{f}_d = \frac{\lambda_1}{\frac{h}{z_d} - \lambda_1} \begin{bmatrix} u_t - u_0 \\ v_t - v_0 \end{bmatrix}. \quad (21)$$

Furthermore, given that the steering angle δ_f is negligible, (20) can be simplified as follows:

$$\mathbf{f}_v = \frac{v_r \lambda_1}{h} \begin{bmatrix} u_t - u_0 \\ v_t - v_0 \end{bmatrix}. \quad (22)$$

Moreover, when there are no centrifugal effects during straight-line motion and the camera's mounting error (the

nonzero roll angle) is negligible, the optical flow can be expressed in its simplest form as follows:

$$\begin{aligned} \mathbf{f}_d &= \frac{z_d}{\frac{hf_y}{v_t - v_0} - z_d} \begin{bmatrix} u_t - u_0 \\ v_t - v_0 \end{bmatrix} \\ \mathbf{f}_v &= \frac{v_r(v_t - v_0)}{hf_y} \begin{bmatrix} u_t - u_0 \\ v_t - v_0 \end{bmatrix}. \end{aligned} \quad (23)$$

Therefore, it can be observed that the vertical optical flow component exhibits a quadratic relationship with respect to the vertical pixel coordinates in the freespace area. Leveraging this insight, we can adopt strategies similar to those used in U-V-Disparity analysis to effectively depict freespace and incorporate additional geometric constraints for other autonomous driving applications.

IV. EXPERIMENTS

This section presents a comprehensive evaluation of our proposed model, including effectiveness validation, quantitative and qualitative assessments, robustness testing, and real-time performance quantification.

As discussed in Section III, the optical flow can be explicitly represented by pixel coordinates and vehicle poses. This representation can be validated through surface fitting. If the optical flow modeling is incorrect, a significant fitting error may occur, which can serve as a criterion for measuring the accuracy of our model.

A. Evaluation Metrics

We use the following four evaluation metrics to quantify the effectiveness of our proposed optical flow models.

1) Average angular error

$$e_A = \frac{1}{N} \sum_{\mathbf{p}} \cos^{-1} \left(\frac{\langle \mathbf{F}(\mathbf{p}), \hat{\mathbf{F}}(\mathbf{p}) \rangle + 1}{\sqrt{(\|\mathbf{F}(\mathbf{p})\|_2 + 1)(\|\hat{\mathbf{F}}(\mathbf{p})\|_2 + 1)}} \right). \quad (24)$$

2) Average end-point error

$$e_E = \frac{1}{N} \sum_{\mathbf{p}} \|\mathbf{F}(\mathbf{p}) - \hat{\mathbf{F}}(\mathbf{p})\|_2. \quad (25)$$

3) Average absolute error of \mathbf{F}_u modeling

$$e_U = \frac{1}{N} \sum_{\mathbf{p}} |\mathbf{F}_u(\mathbf{p}) - \hat{\mathbf{F}}_u(\mathbf{p})|. \quad (26)$$

4) Average absolute error of \mathbf{F}_v modeling

$$e_V = \frac{1}{N} \sum_{\mathbf{p}} |\mathbf{F}_v(\mathbf{p}) - \hat{\mathbf{F}}_v(\mathbf{p})|. \quad (27)$$

Here, \mathbf{F} and $\hat{\mathbf{F}}$ respectively denote the ground-truth and estimated optical flow maps, $\mathbf{F}_{u,v}$ and $\hat{\mathbf{F}}_{u,v}$ respectively represent the ground-truth and estimated horizontal and vertical optical flow maps, \mathbf{p} denotes a pixel in the freespace area, and N represents the total number of valid pixels used for model evaluation.

TABLE I

QUANTITATIVE EXPERIMENTAL RESULTS ON THE CVC12 DATASET

Scenario Type	Vehicle Speed	e_A (Rad)	e_E (Pixel)
Simple	$\times 1$	0.003	0.609
Simple	$\times 4$	0.003	0.704
Complex	$\times 1$	0.089	0.534
Complex	$\times 4$	0.030	0.244

TABLE II

QUANTITATIVE EXPERIMENTAL RESULTS ON THE KITTI DATASET

Scenario Type	e_A (Rad)	e_E (Pixel)	e_U (Pixel)	e_V (Pixel)
Straight-going	0.036	0.921	0.255	0.465
Turning	0.050	1.005	0.599	0.814

B. Datasets

We utilize the following three datasets to quantify the performance of our proposed optical flow models:

- 1) the CVC12 optical flow [26] dataset, which offers dense optical flow ground truth (in pixels) for both the simple and complex synthetic driving scenarios with respect to different vehicle speeds, making it suitable for a comprehensive evaluation of our proposed displacement-based model's validity;
- 2) the KITTI flow 2012 and 2015 [19], [23] datasets, which provide sparse optical flow ground truth (in pixels) captured in real-world driving scenarios, making it suitable for a comprehensive evaluation of our proposed displacement-based model's robustness and effectiveness;
- 3) the CARLA [27] dataset (a synthetic dataset that we created using the CARLA simulator), which provides dense optical flow ground truth (instantaneous velocities of movement) for synthetic driving scenarios, making it suitable for a comprehensive evaluation of our proposed velocity-based model.

C. Model Effectiveness Validation

The qualitative results on the CVC12 dataset are shown in Fig. 3. It can be observed that the modeling error is minimal, and its distribution is relatively uniform, validating the correctness of our proposed displacement-based model. Table I provides the corresponding e_A and e_E results, where e_A is less than 0.1 rad and e_E is less than one pixel. Based on these results, it is evident that our displacement-based optical flow modeling process exhibits high accuracy and robustness, regardless of varying vehicle speeds and diverse driving scenarios.

Furthermore, we use the KITTI flow datasets to evaluate the effectiveness and robustness of our proposed displacement-based model. The qualitative and quantitative experimental results are presented in Fig. 4 and Table II, respectively. It can be observed that our model achieves an e_A of less than 0.05 rad and an e_E of less than one pixel in almost all the scenarios, demonstrating the exceptional performance of our model in real-world scenarios.

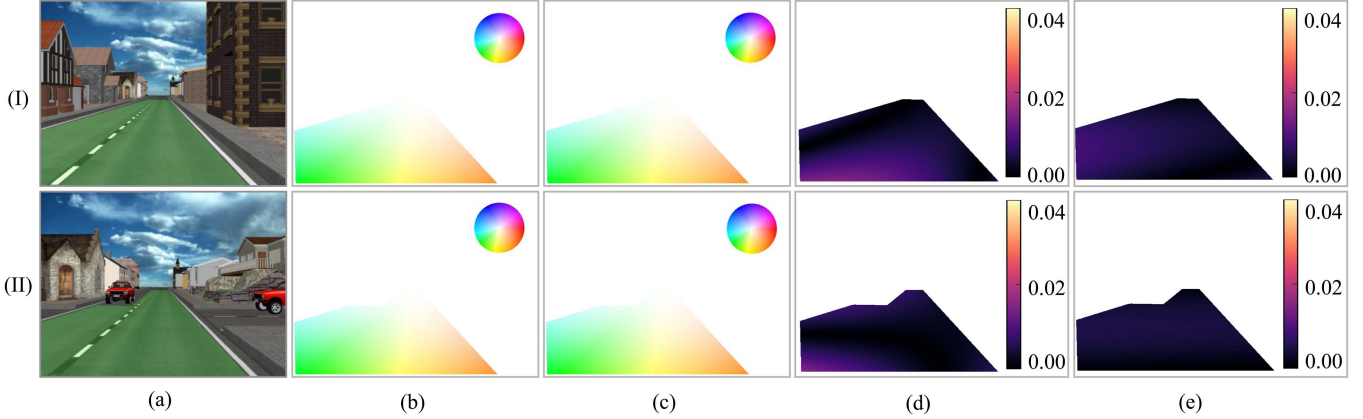


Fig. 3. Qualitative experimental results on the CVC12 dataset: (I) simple scenario and (II) complex scenario. (a) RGB images with freespace shown in green. (b) Optical flow ground truth. (c) Estimated optical flow. (d) Absolute error of F_u modeling. (e) Absolute error of F_v modeling.

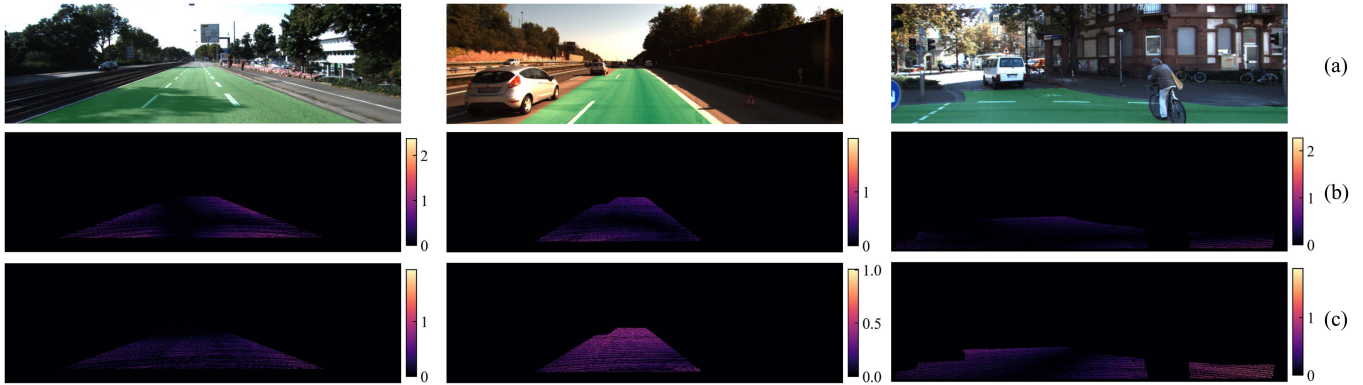


Fig. 4. Qualitative experimental results on the KITTI dataset. (a) RGB images with freespace shown in green. (b) Absolute error of F_u modeling. (c) Absolute error of F_v modeling.

TABLE III
QUANTITATIVE EXPERIMENTAL RESULTS ON THE CARLA DATASET

Driving direction	θ (Degrees)	e_A (Rad/s)	e_E (Pixel/s)
Straight-going	0	0.039	0.526
	5	0.038	0.516
	-5	0.040	0.566
	10	0.034	0.521
	-10	0.041	0.627
Turning	0	0.034	1.260
	5	0.032	1.285
	-5	0.035	1.275
	10	0.033	1.357
	-10	0.037	1.325

Finally, we present the evaluation of our velocity-based models with respect to two types of driving directions and varying camera roll angles on the CARLA dataset. The experimental results are shown in Fig. 5 and Table III. It can be observed that our model achieves high accuracy in θ estimation, with an e_A of less than 0.05 rad/s and an e_E of less than 1.5 pixels/s.

These results strongly indicate the effectiveness of our proposed velocity-based optical flow model.

D. Optical Flow Modeling in the Special Case

We also conduct an additional experiment to demonstrate the robustness of our optical flow modeling process in the special case where $\theta = 0$, $\varphi = 0$, and $x_d = 0$. As shown in Fig. 6, similar to the V-Disparity analysis process, we create the histogram of F_v for each row and then fit the F_v projections to a curve expressed in (23). A fitted F_v map can then be generated using the curve parameters. By comparing the difference between the observed and fitted F_v maps, we obtain an optical flow modeling error of less than one pixel, providing compelling evidence for the validity of our model in this special case.

E. Robustness to Random Noise

To further evaluate the robustness of our proposed model, we conduct an experiment where random Gaussian noises with varying intensities are added to the CVC12 dataset. We then compare the fitting results with the observed optical flow ground truth. As depicted in Fig. 7, our model demonstrates remarkable

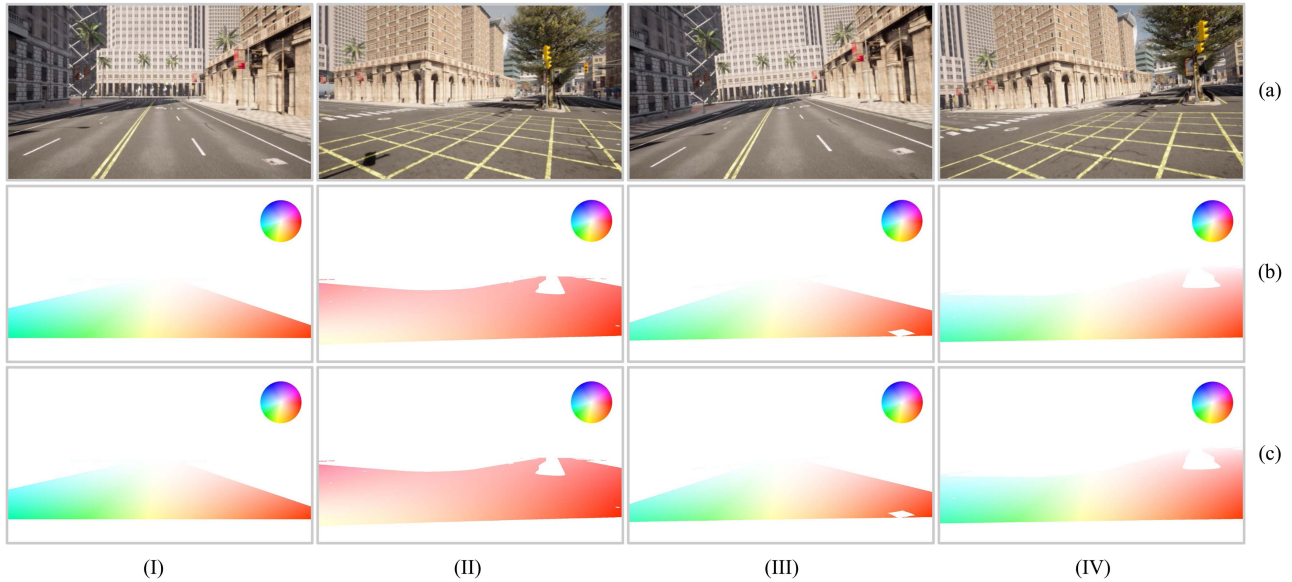


Fig. 5. Qualitative experimental results on the CARLA dataset. (a) RGB images. (b) Optical flow ground truth. (c) Estimated optical flow. (I) Straight-going with $\theta = 0^\circ$. (II) Turning with $\theta = 0^\circ$. (III) Straight-going with $\theta = 5^\circ$. (IV) Turning with $\theta = 5^\circ$.

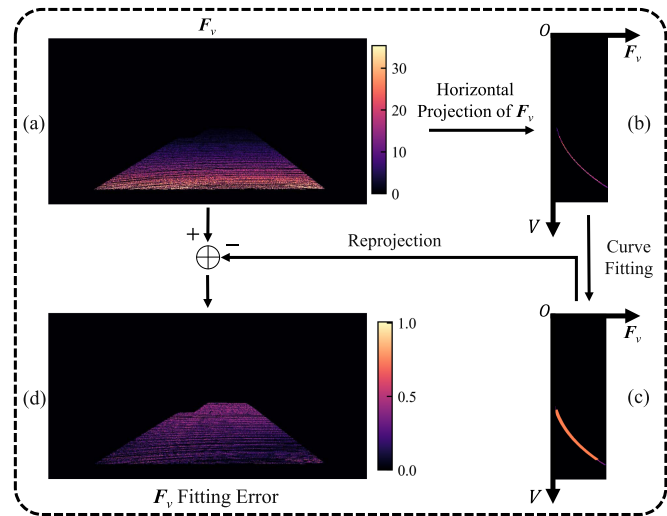


Fig. 6. Optical flow modeling in the special case. (a) F_v . (b) Horizontal projections of F_v . (c) Curve fitting result. (d) F_v fitting error visualization.

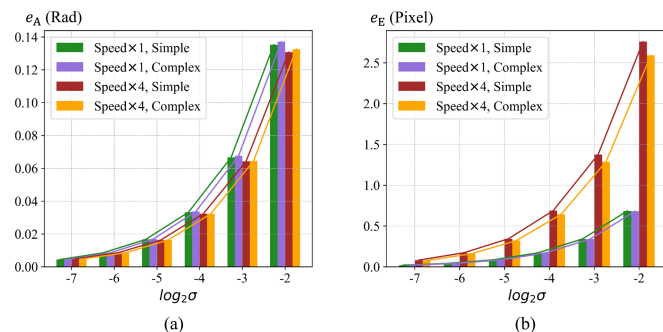


Fig. 7. Noise robustness experiment on the CVC12 dataset. σ represents the variance of added Gaussian noise.

robustness, effectively ignoring noise even when the variance σ is less than $1/32$. In addition, our proposed model exhibits superior performance in terms of end-point error at lower vehicle speeds and maintains consistency across different scenarios. Furthermore, the model demonstrates robustness in terms of the average angular error across various scenarios and vehicle speeds.

F. Real-Time Performance

We conduct an experiment to evaluate the efficiency of our proposed model on an Intel i7-12700K CPU (using only a single thread). With the input of odometry and camera parameters, our model can generate accurate freespace optical flow (resolution: 1242×375 pixels) at a speed of 34.5 frames/s, without any hardware acceleration.

G. Practical Experiments

To validate the feasibility and effectiveness of our proposed freespace optical flow modeling approach, we conduct extensive practical experiments on an automated guided vehicle, as illustrated in Fig. 8. We utilize a well-calibrated stereo rig to collect video sequences in both the indoor and outdoor environments. The optical flow information was obtained using pretrained optical flow networks RAFT [28] (supervised) and ARFlow [29] (unsupervised). We compare the optical flow estimation results without and with our proposed algorithm incorporated. We apply the optical flow to warp the target image to the reference view and calculate the photometric error between the warped target image and the reference image. The results reported in Table IV indicate that our proposed displacement-based model can effectively improve the freespace optical flow results obtained using other deep learning-based algorithms. By applying our modeling approach as a postprocessing step, we are able to improve the geometry

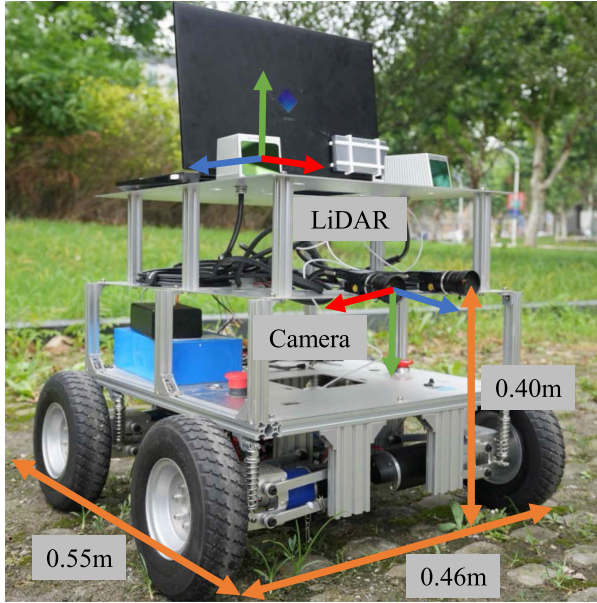


Fig. 8. Illustration of our practical experiments, where all the sensors are well calibrated and synchronized to ensure accurate data collection.

TABLE IV

COMPARISON OF THE PHOTOMETRIC ERROR ACHIEVED BY TWO STATE-OF-THE-ART OPTICAL FLOW ESTIMATION NETWORKS WITHOUT AND WITH OUR PROPOSED ALGORITHM INCORPORATED

Networks	Indoor		Outdoor	
	Baseline	Our model	Baseline	Our model
RAFT [28]	10.423	8.266	9.252	7.399
ARFlow [29]	11.251	9.125	9.953	8.017

consistency of the freespace optical flow. These findings highlight the potential of our approach in enhancing the performance of the existing optical flow algorithms for autonomous driving perception and navigation tasks.

V. POTENTIAL AUTONOMOUS DRIVING APPLICATIONS

A. Freespace Detection

As discussed in Section IV-D, we can leverage our proposed displacement-based optical flow model to segment F_v for freespace detection. Fig. 9 shows an example of freespace detection results achieved using the technique depicted in Fig. 6. This process is akin to the road area detection with disparity segmentation presented in [30].

B. Vehicle Pose Estimation

Our proposed model also holds great promise for vehicle pose estimation by utilizing the information provided by freespace optical flow. As depicted in Fig. 10, we begin by creating a virtual road plane using uniform sampling points. Assuming that the vehicle has a random displacement and yaw angle changes while moving along this virtual road plane, we employ the pinhole camera model to obtain the optical flow ground truth. To evaluate

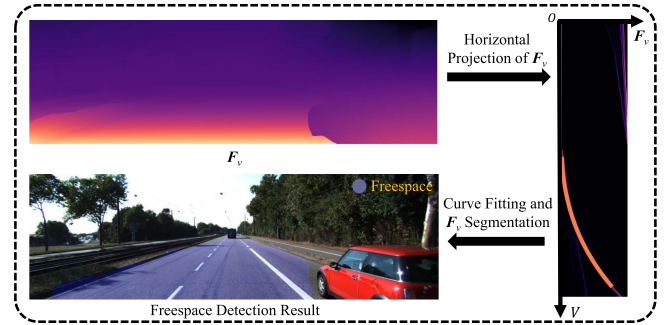


Fig. 9. Freespace detection using our proposed displacement-based optical flow model.

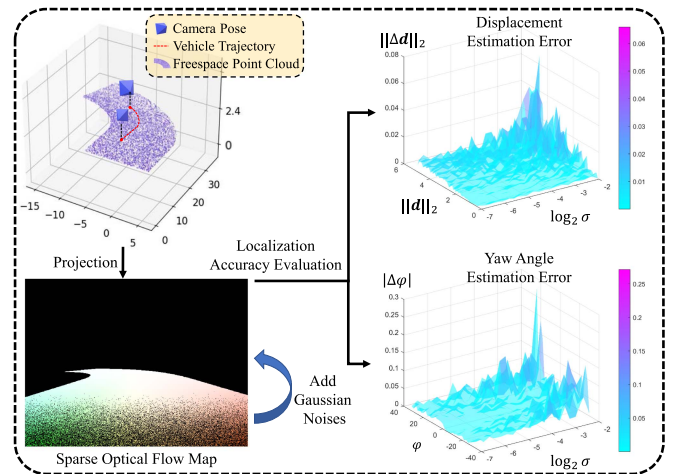


Fig. 10. Vehicle pose estimation with respect to varying displacements, yaw angles, and different levels of Gaussian noise.

the robustness of our model against measurement noise, we introduce Gaussian noise with varying variances σ to the optical flow ground truth, which serves as the sole input to our pose estimation pipeline. We employ the particle swarm optimization algorithm [31] in the bivariate fitting process to determine the best localization parameters. The absolute estimation errors of d and φ are shown on the right. It is evident that the displacement estimation error is below 0.07 m, and the yaw angle estimation error is less than 0.3° . Moreover, the localization accuracy demonstrates robustness against noises of different magnitudes of displacement and rotation, indicating the effectiveness of our approach in accurate pose estimation even under challenging conditions.

VI. CONCLUSION

This article presented two freespace optical flow models for automated driving based on two definitions (displacement and velocity). By leveraging the monocular camera parameters and vehicle poses, we established an explicit relationship between optical flow and the pixel coordinates. We mathematically demonstrated the distribution regularities of both the optical flow models in the PCS. To verify the validity and accuracy of our

proposed models, we conducted extensive experiments on the CVC12 optical flow dataset, KITTI flow 2012 and 2015 datasets, and the CARLA dataset. The average end-point error is less than one pixel on the CVC12 and KITTI datasets and less than 1.5 pixels/s on the CARLA dataset. We also verified the validity of our model in special cases via curve fitting. Our proposed model performed well against noise and demonstrates real-time performance. Finally, we showcased the potential applications of our model through experiments involving freespace detection and vehicle pose estimation.

REFERENCES

- [1] F. Jiménez, J. E. Naranjo, J. J. Anaya, F. García, A. Ponz, and J. M. Armingol, "Advanced driver assistance system for road environments to improve safety and efficiency," *Transp. Res. Procedia*, vol. 14, pp. 2245–2254, 2016.
- [2] N. Kaempchen and K. C. J. Dietmayer, "Data synchronization strategies for multi-sensor fusion," in *Proc. IEEE Conf. Intell. Transp. Syst.*, 2003, vol. 85, no. 1, pp. 1–9.
- [3] R. Fan, H. Wang, P. Cai, and M. Liu, "SNE-RoadSeg: Incorporating surface normal information into semantic segmentation for accurate freespace detection," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 340–356.
- [4] Y. Wu, "Optical flow and motion analysis," *Adv. Comput. Vis. Notes Ser.*, vol. 6, pp. 3–7, 2001.
- [5] M. Zhai, X. Xiang, N. Lv, and X. Kong, "Optical flow and scene flow estimation: A survey," *Pattern Recognit.*, vol. 114, 2021, Art. no. 107861.
- [6] S. Vedula, P. Rander, R. Collins, and T. Kanade, "Three-dimensional scene flow," in *Proc. IEEE Int. Conf. Comput. Vis.*, 1999, vol. 2, pp. 722–729.
- [7] Y. Yu et al., "Accurate and robust visual localization system in large-scale appearance-changing environments," *IEEE/ASME Trans. Mechatronics*, vol. 27, no. 6, pp. 5222–5232, Dec. 2022.
- [8] R. Fan, X. Ai, and N. Dahnoun, "Road surface 3D reconstruction based on dense subpixel disparity map estimation," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 3025–3035, Jun. 2018.
- [9] R. Fan, U. Ozgunalp, B. Hosking, M. Liu, and I. Pitas, "Pothole detection based on disparity transformation and road surface modeling," *IEEE Trans. Image Process.*, vol. 29, pp. 897–908, 2020.
- [10] Z. Hu, F. Lamosa, and K. Uchimura, "A complete U-V-disparity study for stereovision based 3D driving environment analysis," in *Proc. IEEE 5th Int. Conf. 3-D Digit. Imag. Model.*, 2005, pp. 204–211.
- [11] R. Fan, U. Ozgunalp, Y. Wang, M. Liu, and I. Pitas, "Rethinking road surface 3-D reconstruction and pothole detection: From perspective transformation to disparity map segmentation," *IEEE Trans. Cybern.*, vol. 52, no. 7, pp. 5799–5808, Jul. 2022.
- [12] Z. Tu et al., "A survey of variational and CNN-based optical flow techniques," *Signal Processing: Image Commun.*, vol. 72, pp. 9–24, 3 2019.
- [13] B. K. Horn and B. G. Schunck, "Determining optical flow," *Artif. Intell.*, vol. 17, no. 1–3, pp. 185–203, 1981.
- [14] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. 7th Int. Joint Conf. Artif. Intell.*, 1981, vol. 2, pp. 674–679.
- [15] N. Mayer et al., "A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 4040–4048.
- [16] H. Wang, R. Fan, and M. Liu, "CoT-AMFlow: Adaptive modulation network with co-teaching strategy for unsupervised optical flow estimation," in *Proc. Conf. Robot Learn.*, 2021, pp. 143–155.
- [17] X. Shi et al., "FlowFormer++: Masked cost volume autoencoding for pretraining optical flow estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 1599–1610.
- [18] L. Mehl, J. Schmalfluss, A. Jahedi, Y. Nalivayko, and A. Bruhn, "Spring: A high-resolution high-detail dataset and benchmark for scene flow, optical flow and stereo," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 4981–4991.
- [19] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 3354–3361.
- [20] T. Brox, A. Bruhn, N. Papenber, and J. Weickert, "High accuracy optical flow estimation based on a theory of warping," in *Proc. Eur. Conf. Comput. Vis.*, 2004, pp. 25–36.
- [21] A. Dosovitskiy et al., "FlowNet: Learning optical flow with convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 2758–2766.
- [22] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black, "A naturalistic open source movie for optical flow evaluation," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 611–625.
- [23] M. Menze and A. Geiger, "Object scene flow for autonomous vehicles," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3061–3070.
- [24] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski, "A database and evaluation methodology for optical flow," *Int. J. Comput. Vis.*, vol. 92, no. 1, pp. 1–31, 2010.
- [25] J. Y. Wong, *Theory of Ground Vehicles*. Hoboken, NJ, USA: Wiley, 2022.
- [26] N. Onkarappa, "Optical flow in driver assistance systems," Ph.D. dissertation, Dept. Comput. Sci., Univ. Autònoma de Barcelona, Bellaterra, Spain, 2014.
- [27] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," in *Proc. Conf. Robot Learn.*, 2017, pp. 1–16.
- [28] Z. Teed and J. Deng, "RAFT: Recurrent all-pairs field transforms for optical flow," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 402–419.
- [29] L. Liu et al., "Learning by analogy: Reliable supervision from transformations for unsupervised optical flow estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 6489–6498.
- [30] R. Fan and M. Liu, "Road damage detection based on unsupervised disparity map segmentation," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 11, pp. 4906–4911, Nov. 2020.
- [31] D. Wang, D. Tan, and L. Liu, "Particle swarm optimization algorithm: An overview," *Soft. Comput.*, vol. 22, pp. 387–408, 2018.



Yi Feng (Graduate Student Member, IEEE) received the B.E. degree in automation in 2022 from Tongji University, Shanghai, China, where he is currently working toward the M.Sc. degree in control science and engineering with the Machine Intelligence and Autonomous Systems Group, College of Electronics and Information Engineering.

His research interests include computer vision and deep learning.



Ruge Zhang received the B.E. degree in automation from Tongji University, Shanghai, China, in 2023. He is currently working toward the M.Sc. degree in computer science and technology with the High-Performance Computer Research Center, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China.

His research interests include high-performance computing, heterogeneous computing, in-core parallelism, and optimized fast

Fourier transform libraries.



Jiayuan Du received the B.S. degree in automation in 2020 from Tongji University, Shanghai, China, where he is currently working toward the Ph.D. degree in control science and engineering with the Robotics and Artificial Intelligence Laboratory.

His research interests include end-to-end autonomous driving and simultaneous localization and mapping.



Qijun Chen (Senior Member, IEEE) received the B.S. degree in automation from the Huazhong University of Science and Technology, Wuhan, China, in 1987, the M.S. degree in information and control engineering from Xi'an Jiaotong University, Xi'an, China, in 1990, and the Ph.D. degree in control theory and control engineering from Tongji University, Shanghai, China, in 1999.

He is currently a Full Professor with the College of Electronics and Information Engineering, Tongji University. His research interests include robotics control, environmental perception, and understanding of mobile robots and bioinspired control.



Rui Fan (Member, IEEE) received the B.Eng. degree in automation from the Harbin Institute of Technology, Harbin, China, in 2015, and the Ph.D. degree (under the supervision of Prof. John G. Rarity and Prof. Naim Dahnoun) in electrical and electronic engineering from the University of Bristol, Bristol, U.K., in 2018.

From 2018 to 2020, he was a Research Associate (under the supervision of Prof. Ming Liu) with the Hong Kong University of Science and Technology, Hong Kong. From 2020 to 2021, he was a Postdoctoral Scholar-Employee (under the supervision of Prof. Linda M. Zangwill and Prof. David J. Kriegma) with the University of California San Diego, La Jolla, CA, USA. In 2021, he joined the College of Electronics and Information Engineering, Tongji University, Shanghai, China, as a Full Research Professor. In 2022, he became a Full Professor with the College of Electronics and Information Engineering, Tongji University, as well as with the Shanghai Research Institute for Intelligent Autonomous Systems. His research interests include computer vision, deep learning, and robotics.

Dr. Fan was an Associate Editor for 2023 ICRA and 2023 IROS, and a Senior Program Committee Member of 2023 and 2024 AAAI. He is the General Chair of the AVVision community and organized several impactful workshops and special sessions in conjunction with 2021 WACV, 2021/2022/2023 ICIP, 2021 IEEE ICCV, and 2022 ECCV. He was named in Stanford University List of Top 2% Scientists Worldwide in 2022.