

Direct Imaging of Stabilized Optical Flow and Possible Anomalies From Moving Vehicle

Shigeru Ando¹, *Member, IEEE*, and Toshiki Kindo, *Member, IEEE*

Abstract—Machine perception of dynamic scenes becomes more and more important for autonomous vehicles and vision-based driver-assistance systems. Even with other 3D ranging devices, dense, detailed and instantaneous detection of optical flow is essential for early distinguishing small moving objects in the 3D environment from the moving vehicle. To overcome the limited performance in the immediacy, resolution, accuracy and acuity of existing methods, we provide an optical flow detection scheme based on a three-phase correlation image sensor (3PCIS) that is capable of Fourier-coefficient imaging combined with an exact and direct algorithm derived from the weighted integral method of identifying the differential equation model from a short-duration observation. To utilize inherent performances of the detection scheme by removing the large and rapid disturbances induced by the rotational fluctuations of the platform, we introduce a software operation of gaze in which the image coordinates are fixed on and smoothly pursue a forward stable object so that the optical flow field is relative to the moving coordinate system. In it, the gaze subsystem continuously provides angular velocity and pose between the camera and gaze target, while the imaging subsystem instantaneously obtains two optical flow distributions by cancelling the ego-rotation components and then removing the outwardly diverging components derived mainly from stationary 3D environments. Possible anomalies captured in each frame instantaneously provide candidates of hazardous objects that should be tracked and further investigated. We examine the performance of optical flow stabilization and anomaly detection using image sequences of monocular 3PCIS mounted on a moving vehicle on town roads and a highway.

Index Terms—Optical flow, ego-motion, gaze, autonomous vehicle, weighted integral method, correlation image sensor.

I. INTRODUCTION

THE early detection of harmful traffic situations is an important subject for autonomous vehicles and vision-based driver-assistance systems navigating around other moving vehicles and humans. Future vehicles must be able to perceive their environment reliably from their own visual input, just like a careful human driver with keen senses. In DARPA Grand/Urban Challenges for practical autonomous driving technologies [1], [2], LiDAR is a main device for capturing 3D traffic environments [3]. However, its resolution

is too low to detect an early sign of hazard for planning an emergent reaction to avoid it [4], [5]. The role of vision sensors remains important in the dense and rapid detection of anomalous conditions such as a ball and children jumping into a road, a motorcycle approaching an intersection without slowing down, and a vehicle that changes lanes without considering the surroundings. Toward this aim, dense optical flow (OF) is particularly advantageous since it not only instantaneously and widely captures the environment and objects, but also provides insights into their geometric layout that can be easily fused with LiDAR or millimeter-wave radars. However, under the unknown ego-motion of the advancing platform, the tasks become considerably difficult for detecting small moving objects in cluttered backgrounds [6].

The methods for OF detection [7], [8] comprise 1) a description of brightness constancy during motion, 2) the local or global modeling of the velocity field, and 3) an optimized solution with regularization, and they all have long been studied. The techniques for 2) and 3) are mostly general frameworks, but those for 1) are highly specific to optical images. The most established one is the optical flow partial differential equation (OFPDE), which describes strictly the spatiotemporal variation of image intensity over the velocity field. However, the OFPDE in the differential methods (DMs) [9], [10] is followed immediately by a dirty approximation of the temporal differential using the consecutive-frame difference and thus suffers a fundamental limitation in the range of velocity and accuracy. Also, an OFPDE can constrain only a normal component of the flow. Thus, the aims of many studies have been the regularization of these problems, such as locally smooth [10] or globally coherent flows [11], multiscale analysis [12] and extended constancy assumptions [13]. Another category of 1) is the multiframe correspondences of sparse features in an image sequence [14], [15]. Although it provides long-term reliable results, an obvious disadvantage is that small components without distinctive features can be ignored. To implement an on-vehicle early warning vision under ego-motion, two different roles of the OF must be coupled and enhanced together. The tracking determines both the velocities and loci of featured points, which enable accurate recoveries of ego-motion and the related OF. Then, the OFPDE, with proper mathematical treatments, provides a stable, detailed and instantaneous description of the flow field.

Recently, a novel framework for these studies has been obtained as a combination of a mathematical technique, the weighted integral method (WIM), and an imaging device, the three-phase correlation image sensor (3PCIS) [16]. The WIM

Manuscript received 26 October 2021; revised 16 May 2022 and 9 August 2022; accepted 12 August 2022. Date of publication 26 August 2022; date of current version 5 December 2022. The Associate Editor for this article was S. S. Nedevschi. (Corresponding author: Shigeru Ando.)

Shigeru Ando is with the Department of Information Physics and Computing, The University of Tokyo, Tokyo 113-8656, Japan (e-mail: shigeru_ando@ipc.i.u-tokyo.ac.jp).

Toshiki Kindo is with the Kanazawa Institute of Technology, Kanazawa 921-8501, Japan.

Digital Object Identifier 10.1109/TITS.2022.3199203

provides an exact algebraic equation (AE) for determining the coefficients of a differential equation from weighted integral measurements of its variables [17], [18], [19], [20], [21]. For the OFPDE, an exact closed-form solution of the velocity is obtained from short-time Fourier coefficients (FCs) of time-varying intensity [22], [23] captured directly by the 3PCIS. Various types of 3PCIS have been fabricated [24], [25] and improved for an accurate algebraic solution of OF [26].

The purpose and the main contribution of this paper is to present an on-vehicle OF detection scheme with an extremely low latency and an extended spatiotemporal resolution free of the accuracy problems and ill-conditions in the traditional methods. The most severe impediments to obtain practicable performances are the rapidly time-varying OF components induced by rotational fluctuations of the platform. To remove them, we introduce a software operation of gaze driven by the instantaneous image motion, in which the image coordinate is fixed on a stable stationary object. Thus, it pursues it so that the OF field is relative to the moving coordinate. Owing to the extended performances of the WIM and 3PCIS, both the image coordinate and the relative OF fields are stabilized up to their time-differential order such that only the translational ego-motion components and flows due to moving objects (if they exist) remain.

The method we propose is for a realtime, direct imaging of anomalous objects as inputs to succeeding high-level analysis; thus, it is restricted to an early, pre-attentive (near-sensor) level. The procedure is based solely on the mathematical models, dominant statistics and algebraic operations. Also, it is so designed to permit feedback from the cognitive level in the integration. In Section III, the WIM for ego-motion OF is described. Then, in Section IV, the stable visual coordinate (SVC) and gazeflow are introduced. In Section V, we describe the algorithms in detail, while in Section VI, we examine them using real 3PCIS image data captured from a vehicle driven on busy town roads and on a highway.

II. RELATED WORK

A. Ego-Motion and Moving-Object Detection in 3D Scenes

The ego-motion-induced OF in a static 3D scene [27] was implemented with the DM for detecting rotation-only or translation-only ego-motion [28], for the focus of expansion (FOE) on a rigid object [11], and for the robust 3D reconstruction [29]. The problems encountered when moving objects are present have been studied as the structure from motion in a multibody environment. For separating static scenes and rigidly moving objects, most methods are based on the random sample consensus (RANSAC) framework and require a significant number of iterations [30]. Further considerations include the use of the factorization method [14], a piecewise planar model of a 3D scene [15], probabilistic reasoning including occlusion [31] etc. The visual odometry or simultaneous localization and mapping (SLAM) [33] provide a reliable framework for visual navigation. Applications to vehicles include the estimation of ego-motion using the road surface [34], the detection of position/velocity of other vehicles [35] and the inference of road layout for autonomous

driving [36]. Additional considerations include tight coupling with multiperson detection [37], the use of vehicle kinematics constraint [38], robust circular matching in tracking and stereo [39], context-aware motion descriptor using oriented histograms of OF [40] etc. However, in these methods, the instantaneous detection of small moving objects for traffic safety is not a primal concern.

B. Motion Anomaly Detection From a Moving Vehicle

For vehicle safety, a child, bicycle or ball approaching the road represent targets of interest that requires an emergent reaction. The difference between ego-motion-compensated frames of stereo cameras [41], [42], [43] or a monocular camera [44], [45] provides the instantaneous distribution of possible anomalies. The frame difference, however, includes various noises such as occluding and/or occluded edges or misaligned textured regions. A reliable operation for rejecting them is the continuous tracking and detection described in the previous section as well as the semantic inference of moving targets [46], detection and tracking via composite object description [47], fusing OF orientation and magnitude for robust obstacle detection [48], use of a learning framework for robustness [49] and maintaining occupancy probability in voxels using 3D point cloud and odometry [50], which often reduce the rapidness of anomaly detection and sensitivity to small objects. Real-time approaches with binocular stereo and other sensors include pedestrian detection from stereo depth histograms [51], moving object detection using FOE clustering and inertial sensors [52] and obstacle detection by fusing stereo and radar [53]. Among others, the direct and single-frame detection of ego-motion-compensated OF with an extended reliability is desirable for removing erroneous responses and solving the spatiotemporal resolution problems simultaneously.

C. Studies Inspired by the Human Vision

The roles of gaze or fixation in human and machine vision have been indicated in various studies on the following topics: recovering ego-motion [54], geometrical reasoning of object shape [55] and facilitating obstacle avoidance and target following [56]. The ultrashort latencies of the conjugate eye responses to ego-motion and pursuing a moving object are shown to be enabled by feedforward and feedback mechanisms based on OF [57]. An active vision system using a pan/tilt platform [58] is a hardware realization of gaze. In contrast, software realization is less costly using a coordinate transform and geometrical correction on the image whenever the gaze target is present in the image area. Other studies include the foveated vision with space-variant resolution [59], an event camera with time-resolved sensitivity [60] and a small moving object detector inspired by the fly visual system [61].

III. OPTICAL FLOW UNDER EGO-MOTION

A. Optical Flow Differential Equation for Ego-Motion

Let $\mathbf{r} = (x, y)$ and (X, Y, Z) be the image and camera coordinates fixed on the vehicle (see Fig. 1). Let the rotation

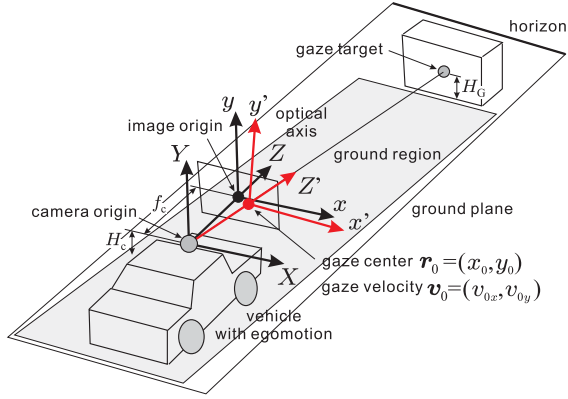


Fig. 1. Definitions of coordinate systems. The camera coordinates (X, Y, Z) and image coordinates (x, y) are fixed on the moving vehicle. The gaze target is a stationary object in the advancing direction of the vehicle. The stable visual coordinates are (x', y', Z') (oblique coordinates), in which x' and y' are perpendicular and in the image plane, x' is parallel to the horizon, and Z' is the line of sight toward the gaze target.

and translation motions of the camera be $\Omega = (\Omega_x, \Omega_y, \Omega_z)$ and $T = (T_x, T_y, T_z)$, respectively. Then, for the stationary object surfaces free from occlusions, the OF field from the moving camera is expressed as [27], [29]

$$\mathbf{v}(\mathbf{r}) = \mathbf{B}(\mathbf{r})\Omega + \frac{1}{Z(\mathbf{r})}\mathbf{A}(\mathbf{r})\mathbf{T}, \quad (1)$$

where $Z(\mathbf{r})$ is the distance to the surface being imaged at \mathbf{r} . The 2×3 matrices $\mathbf{B}(\mathbf{r})$ and $\mathbf{A}(\mathbf{r})$ are expressed as

$$\mathbf{B}(\mathbf{r}) = \begin{bmatrix} xy/f_c & -f_c - x^2/f_c & y \\ f_c + y^2/f_c & -xy/f_c & -x \end{bmatrix} \quad (2)$$

$$\mathbf{A}(\mathbf{r}) = \begin{bmatrix} -f_c & 0 & x \\ 0 & -f_c & y \end{bmatrix}, \quad (3)$$

where f_c is the focal length of the camera. By using Eq. (1), the OFPDE induced by ego-motion is expressed as

$$\left\{ (\mathbf{B}(\mathbf{r})\Omega + \frac{1}{Z(\mathbf{r})}\mathbf{A}(\mathbf{r})\mathbf{T}) \cdot \nabla + \frac{\partial}{\partial t} \right\} f(\mathbf{r}, t) = 0, \quad (4)$$

where $f(\mathbf{r}, t)$ is the light intensity on the image plane.

B. Exact Short-Time Integral Form of OFPDE

Assume Eq. (4) is satisfied in the frame interval $[0, T]$ of the camera and the temporal changes of $\mathbf{v}(\mathbf{r})$ in the frame are ignorable in comparison with those of $f(\mathbf{r}, t)$. Then, the OFPDE in $[0, T]$ is identically expressed by optical flow algebraic equations (OFAEs, see [26] for derivation):

$$\begin{aligned} & (\mathbf{B}(\mathbf{r})\Omega + \frac{1}{Z(\mathbf{r})}\mathbf{A}(\mathbf{r})\mathbf{T}) \cdot \nabla g_n(\mathbf{r}) \\ & + [f(\mathbf{r}, t)]_0^T + jn\Delta_\omega g_n(\mathbf{r}) = 0 \quad \forall n, \end{aligned} \quad (5)$$

where $\Delta_\omega \equiv 2\pi/T$ is the unit frequency, and

$$g_n(\mathbf{r}) = \int_0^T f(\mathbf{r}, t) e^{-jn\Delta_\omega t} dt \quad (6)$$

is the n th order Fourier coefficient in $[0, T]$ which can be directly captured by the 3PCIS ($n = 0$ is the intensity image

TABLE I
COMPARISONS OF OPTICAL FLOW DETECTION BASED ON THE WIM WITH 3PCIS AND THE DM WITH A CONVENTIONAL IMAGE SENSOR (IS)

item	WIM + 3PCIS	DM + IS
basic equation	exact OFAE	numerical OFPDE
frames used	1 (single exposure)	2 (consecutive)
images used	0th Fourier coef. $g_0(\mathbf{r})$ 1st Fourier coef. $g_1(\mathbf{r})$	intensity $f(\mathbf{r}, t)$ intensity $f(\mathbf{r}, t+T)$
velocity field	constant in a frame	constant in two frames constant locally
per-pixel solution	2D velocity	1D normal velocity
main noise	shot (photon) noise	approximation error
velocity range	no theoretical limit	a few pixels/frame

captured always). By eliminating the integral boundary term $[f(\mathbf{r}, t)]_0^T$ using $n = 0$ and $n = 1$, the OFAE is expressed as

$$\begin{aligned} & (\mathbf{B}(\mathbf{r})\Omega + \frac{1}{Z(\mathbf{r})}\mathbf{A}(\mathbf{r})\mathbf{T}) \cdot \nabla (g_1(\mathbf{r}) - g_0(\mathbf{r})) \\ & + j\Delta_\omega g_1(\mathbf{r}) = 0. \end{aligned} \quad (7)$$

For the whole image, the system of Eq. (7) involves six unknowns of Ω and \mathbf{T} plus pixelwise unknowns of $Z(\mathbf{r})$. In the same time, Eq. (7) is complex, providing two equations pixelwise. Therefore, in the WIM, all unknowns are essentially solvable except for a common scale of \mathbf{T} and $Z(\mathbf{r})$ if the 3D environment is stationary (no moving objects). It is not the case for the DMs using Eq. (4) which provides a real equation pixelwise.

The differences in mathematical backgrounds and performance for OF detection between the WIM using the OFAE and a 3PCIS, on one side, and the DM using the numerical approximation of OFPDE and a conventional image sensor, on the other side, are summarized in Table I. For experimental comparisons between them and for details of the 3PCIS, see Section VI-A and [26].

IV. STABLE VISUAL COORDINATES AND GAZEFLOW

A. Definitions

Gaze defined in this paper is the operation of keeping the view center and horizontal axis of the image plane at the images of a steady object ahead (gaze target) and the horizontal line in the environment, respectively. For the geometry, see Figs. 1 and 5 and their captions. The view center and horizontal axis make up a moving coordinate system in the image plane. Stability of OF is, therefore, achieved in both egocentric and object-centric senses because the origin is at the moving vehicle, while the angles of axes are determined in relation to the gaze target and the horizon fixed in the world coordinate. A stable visual coordinate (SVC) is the frame determined by the line of sight connecting the camera origin and the point on the gaze target (gaze center) as the Z' axis, and the moving coordinates in the image plane as the x' and y' axes. Gaze-flow is the OF defined in the SVC. The operation to acquire the stable visual coordinate is the gaze. At the gaze center, the gaze-flow must be always zero since the SVC continuously and smoothly pursues the gaze target. The gaze-flow describes a detailed motion distribution relative to the gaze center. This is an important difference of the

gaze-flow from the OF stabilized by a gyro sensor (see Fig. 8 in Section V for the difference between the gaze parameters and gyro-sensor outputs).

B. Gaze-flow: Optical Flow Under Gaze Operation

Let the position of the gaze center be $\mathbf{r}_0 = (x_0, y_0)$. Let the distance to the gaze target be $Z(\mathbf{r}_0) \equiv Z_G$. The OF at \mathbf{r}_0 is expressed as

$$\mathbf{v}_0 = \mathbf{B}(\mathbf{r}_0)\boldsymbol{\Omega} + \frac{\mathbf{A}(\mathbf{r}_0)\mathbf{T}}{Z_G}. \quad (8)$$

We also express \mathbf{r}_0 and \mathbf{v}_0 as smooth temporal functions $\mathbf{v}_0(t) = \dot{\mathbf{r}}_0(t)$ for the inter-frame Kalman filtering. The method of obtaining \mathbf{r}_0 , \mathbf{v}_0 and $\boldsymbol{\Omega}_z$ is described in Section V.

Now, let us consider a velocity distribution being relative to the gaze center as

$$\mathbf{v}(\mathbf{r}) - \mathbf{v}_0 = (\mathbf{B}(\mathbf{r}) - \mathbf{B}(\mathbf{r}_0))\boldsymbol{\Omega} + \left(\frac{\mathbf{A}(\mathbf{r})}{Z(\mathbf{r})} - \frac{\mathbf{A}(\mathbf{r}_0)}{Z_G} \right)\mathbf{T}, \quad (9)$$

in which the rotational motion term is expressed as

$$\begin{aligned} & (\mathbf{B}(\mathbf{r}) - \mathbf{B}(\mathbf{r}_0))\boldsymbol{\Omega} \\ &= \begin{bmatrix} y - y_0 \\ -x + x_0 \end{bmatrix} \boldsymbol{\Omega}_z + \frac{1}{f_c} \begin{bmatrix} xy - x_0y_0 & -x^2 + x_0^2 \\ y^2 - y_0^2 & -xy + x_0y_0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\Omega}_x \\ \boldsymbol{\Omega}_y \end{bmatrix}. \end{aligned} \quad (10)$$

For the second-order small term on the right-hand side, let us introduce the approximations that the gaze center is sufficiently near the optical center, i.e., $\|\mathbf{r}_0\| \ll f_c$, and the gaze target is sufficiently distant, thus $\|(T_x, T_y)\|/Z_G \ll \|(\boldsymbol{\Omega}_x, \boldsymbol{\Omega}_y)\|$. Then, the small term is expressed as

$$\frac{1}{f_c^2} \begin{bmatrix} x^2 - x_0^2 & xy - x_0y_0 \\ xy - x_0y_0 & y^2 - y_0^2 \end{bmatrix} \mathbf{v}_0 = \frac{\mathbf{r}\mathbf{r}^t - \mathbf{r}_0\mathbf{r}_0^t}{f_c^2} \mathbf{v}_0 \quad (11)$$

using the gaze center velocity \mathbf{v}_0 . Therefore, by subtracting from $\mathbf{v}(\mathbf{r})$ the velocity distribution determined by \mathbf{r}_0 , \mathbf{v}_0 and $\boldsymbol{\Omega}_z$, the gaze-flow distribution is expressed as

$$\begin{aligned} \mathbf{v}_G(\mathbf{r}) &\equiv \mathbf{v}(\mathbf{r}) - \mathbf{v}_0 - \frac{\mathbf{r}\mathbf{r}^t - \mathbf{r}_0\mathbf{r}_0^t}{f_c^2} \mathbf{v}_0 - (\mathbf{r} - \mathbf{r}_0)^\perp \boldsymbol{\Omega}_z \\ &= \left(\frac{f_c}{Z_G} - \frac{f_c}{Z(\mathbf{r})} \right) \begin{bmatrix} T_x - x_0T_z/f_c \\ T_y - y_0T_z/f_c \end{bmatrix} + \frac{T_z}{Z(\mathbf{r})}(\mathbf{r} - \mathbf{r}_0), \end{aligned} \quad (12)$$

where \perp indicates the $\pi/2$ rotation of the vector. The gaze-flow $\mathbf{v}_G(\mathbf{r})$ is the sum of a unidirectional vector field of the first term and a diverging vector field of the second term. The first term vanishes where the distance is near the gaze target. In this respect, gaze is the procedure of making the description of the OF field simpler by excluding the effects of $\boldsymbol{\Omega}$ everywhere and those of T_x and T_y near the distance of the gaze target.

To understand the terms $(x_0, y_0)T_z/f_c$ subtracted from (T_x, T_y) in Eq. (12), see Fig. 2. The direction of the line of sight (Z' axis) and the velocity toward it are (θ_x, θ_y) and T'_z , respectively. Then, if $\|\mathbf{r}_0\| \ll f$, the velocity components induced by $T'_z \simeq T_z$ along the x and y axes are

$$T'_z \sin \theta_x \simeq \frac{y_0 T_z}{f_c}, \quad -T'_z \sin \theta_y \simeq \frac{x_0 T_z}{f_c}, \quad (13)$$

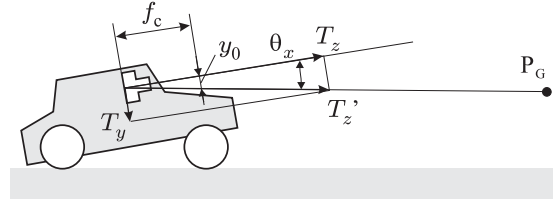


Fig. 2. Translational ego-motion components T_x and T_y induced by the forward motion T'_z and the pose rotation (side view for T_y is shown). Gaze operation is also capable to cancel these components from OF.

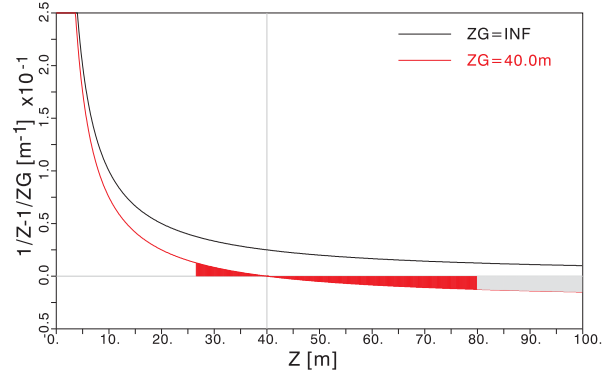


Fig. 3. Plots of $1/Z - 1/Z_G$ vs Z when $Z_G = 40\text{m}$ and $Z_G = \infty$. The function value is proportional to the influence of T_x and T_y in the gaze-flow, and it is zero when $Z = Z_G$. For finite Z_G , the influence becomes smaller in $Z < 2Z_G$ than when $Z_G = \infty$. It is bounded within the widest range of Z (e.g., red zone) when Z_G is the harmonic mean of the terminal values.

which are equal to the subtracted terms. This means that the components T_x and T_y induced by the tilt of the optical axis from the line of sight have been removed in the gaze-flow.

C. Desired Conditions of the Gaze Target

The gaze target should be, for reliable OF detection, richly patterned, occlusion-free, and stationary during the vehicle advance. Thus, it is in the forward direction. Regarding its distance, Fig. 3 shows graphs of the residual influence of T_x and T_y in the gaze-flow. By choosing a finite gaze-target distance Z_G , the influence of T_x and T_y becomes far smaller than in the case of $Z_G = \infty$. Generally, the influence is bounded equally at extremes of a distance range $Z_{\min} < Z < Z_{\max}$ when $Z_G = 2/(1/Z_{\min} + 1/Z_{\max})$. If $Z_{\max} = \infty$ on the far side, $Z_{\min} = Z_G/2$ on the near side. In practice, the distance is unknown. Hence, an appropriate algorithm is designed to choose the gaze target as a distant object with a large area (see Section V).

On the relationship to the ground plane, let the heights of the target and camera be H_G and H_c , respectively. Then, the image height in the SVC of the ground plane at the distance $Z(\mathbf{r})$ is expressed as

$$y' \simeq -\frac{f_c}{Z(\mathbf{r})}H_c + \frac{f_c}{Z_G}(H_c - H_G), \quad (14)$$

and if $H_c = H_G$, the second term including Z_G (decreasing with locomotion) vanishes. Therefore, by adopting a gaze target at the same height as the camera, the ground plane image

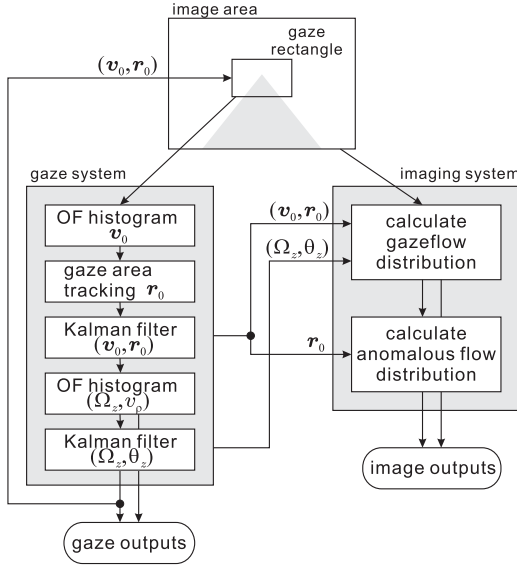


Fig. 4. Block diagram of the overall algorithm. The gaze system maintains the gaze area and continues tracking or initializing it. The imaging system, using the gaze system results, performs the imaging operation of gazeflow and anomalous flow by cancelling the ego-rotation components of the OF and tangential and positive divergent components of the gazeflow, respectively.

has a simple and stationary distance distribution. This is the case even after changing the gaze target.

V. PRACTICAL MATTERS AND ALGORITHMS

The algorithms for determining the gazeflow comprise 1) pursuing operation of the gaze target to determine the origin and velocity $(\mathbf{r}_0, \mathbf{v}_0, \Omega_z)$ of the SVC, and 2) detecting the gazeflow as an instantaneous OF field free from the above-determined ego-motion components. Note that the strategies of 1) and 2) should be contrastive: 1) requires a reliable steady object selection, smoothness, accuracy and long-time consistency as the ego-motion of the platform, while 2) requires immediateness and spatio-temporal resolution as an early warning mechanism in a possibly harmful environment. The algorithms are integrated together as shown in Fig. 4.

A. Placement of Gaze Area and Velocity Estimate

The gaze rectangle is, as shown in Fig. 5, an extended region (128×72 pixels area in 704×512 pixels image) attached to the gaze center for velocity estimation and image matching. The determination of the gaze center is activated initially, when the offset of the gaze center from the optical axis exceeds a threshold or when tracking fails. The center is placed at a local maximum of image variance near the optical center. Across a saccadic change of the gaze center, the gaze parameters \mathbf{v}_0, Ω_z and θ_z are transferred and \mathbf{r}_0 is shifted accordingly.

The gaze area is a subset of the gaze rectangle consisting solely of the target image and is determined based on the clustering of OF. First, the velocity distribution is obtained using the local least-squares method expressed as

$$\mathbf{v}(\mathbf{r}) = \arg \min_{\mathbf{v}} \iint_{\Gamma(\mathbf{r})} |\mathbf{v} \cdot \nabla(g_1 - g_0) + j \Delta_{\omega} g_1|^2 d\mathbf{r}, \quad (15)$$

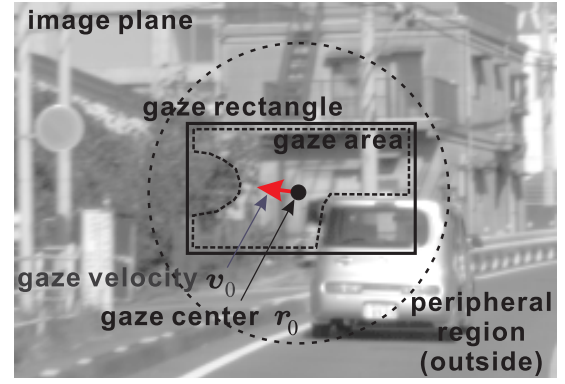


Fig. 5. Definitions of the gaze center, gaze rectangle and gaze area. The gaze area is the largest subarea of the gaze rectangle with a uniform OF. The gaze center is attached to a 2D-featured pattern in the gaze area. The peripheral region is used to estimate the rotational and radial OF components.

where $\Gamma(\mathbf{r})$ is a small area (e.g, 3×3 pixels) around \mathbf{r} . The normal equation to obtain $\mathbf{v} = (v_x, v_y)$ is expressed as

$$\Re \left\{ \iint_{\Gamma(\mathbf{r})} \nabla(g_1 - g_0)^* (\nabla(g_1 - g_0))^t d\mathbf{r} \begin{bmatrix} v_x \\ v_y \end{bmatrix} \right\} = \Delta_{\omega} \Im \left\{ \iint_{\Gamma(\mathbf{r})} \nabla(g_1 - g_0)^* g_1 d\mathbf{r} \right\}, \quad (16)$$

where $\Re\{\cdot\}$ and $\Im\{\cdot\}$ are the real and imaginary parts, respectively. From it, a 2D histogram of $\mathbf{v}(\mathbf{r})$ is generated in the gaze rectangle. While weighting the similarity with the previous velocity $\mathbf{v}_0(t - T)$, the peak is selected as $\mathbf{v}_0(t)$. The variance of scatter around the peak is used to extract the gaze area using the Mahalanobis distance of $\mathbf{v}(\mathbf{r})$ from $\mathbf{v}_0(t)$.

By using the peak velocity as $\mathbf{v}_0(t)$, the largest object with a uniform OF is chosen as the gaze target. Objects with spatially varying OF, such as the road surface, are excluded. The selection of the rear face of the preceding vehicle is usually not excluded since it has little effect on visual stability. Also, a large crossing object near the path can be gazed. For a moment, the gazeflow becomes relative to the object to describe its detailed motion (see Fig. 14(d) for an example). Such gaze targets are reset whenever the condition becomes unacceptable for stable tracking.

B. Tracking of the Gaze Area With Enhanced Resolution

Images from a moving vehicle suffer motion blur, which reduces the matching accuracy. With the 3PCIS, blur can be reduced by combining the intensity and correlation images. A truncated Fourier series expansion of the time function $f(\mathbf{r}, t)$ of the incident light with $g_{-1}(\mathbf{r}) = g_1^*(\mathbf{r})$, $g_0(\mathbf{r})$ and $g_1(\mathbf{r})$ is expressed as

$$f(\mathbf{r}, t) \simeq g_0(\mathbf{r}) + 2\Re\{g_1(\mathbf{r})e^{j\Delta_{\omega}t}\} \quad (0 < t < T), \quad (17)$$

which leads to the central-time image (CTI)

$$f^{[m]}(\mathbf{r}) \equiv f\left(\mathbf{r}, \frac{T}{2}\right) \simeq g_0(\mathbf{r}) - 2\Re g_1(\mathbf{r}), \quad (18)$$

where the superscript m is the frame count. The first spectral zero due to the motion blur is removed [22], [23], and spatial bandwidth of the CTI is increased to about three times $g_0(\mathbf{r})$.

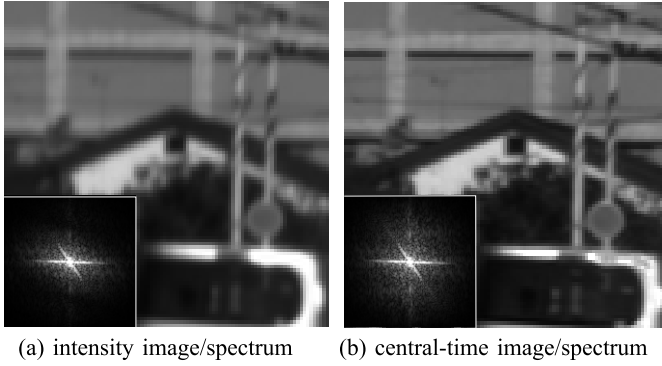


Fig. 6. Motion blur reduction using intensity and correlation images (a part of Fig. 12(c)). The Fourier series expansion at the central time of the frame provides a removal of first spectral zeros of blur and a resolution enhancement about three times in the motion direction (upper right).

Fig. 6 shows an example of motion blur reduction. Resolution enhancement in the motion direction (mostly by pitch and yaw) is achieved without any false effects such as ringing at edges. In the following matching operation, the CTIs are used to increase their selectivity and accuracy.

Even in the gaze area, different appearances of the gaze target can occur. Thus, the matching criterion is the maximization of the sum of similarities being tolerant of the inclusion of unmatched pixels defined as

$$\mathbf{r}_0^{[m]} = \arg \max_{\mathbf{r}_0^{[m]}} \int \int_{\text{gaze area}} \exp \left\{ -\frac{1}{2\sigma^2} (f^{[m]}(\mathbf{r} - \mathbf{r}_0^{[m]}) - f^{[m-1]}(\mathbf{r} - \mathbf{r}_0^{[m-1]}))^2 \right\} d\mathbf{r}. \quad (19)$$

where σ specifies an acceptable deviation for the similarities. Since the gaze area is small, the expansion and rotation of the previous image in it are ignorable. The initial position of the search is at $\mathbf{r}_0^{[m-1]} + \mathbf{v}_0^{[m]}$. The subpixel offset is then determined by the differential method.

C. Estimation of Ω_z

The roll velocity Ω_z is determined using $(\mathbf{v}_0, \mathbf{r}_0)$ after Kalman filtering. Peripheral regions (see Fig. 5) are more informative for this estimate. To exclude small moving objects and estimate Ω_z in stationary regions, we again use the histogram method. Let $v_\rho(\mathbf{r}) \equiv T_z/Z(\mathbf{r})$ be the radial velocity coefficient. Then, the local least-squares method to determine the rotational and radial velocity coefficients is expressed as

$$\begin{aligned} & (\Omega_z(\mathbf{r}), v_\rho(\mathbf{r})) \\ &= \arg \min_{\Omega_z, v_\rho} \int \int_{\Gamma(\mathbf{r})} |(\Omega_z(\mathbf{r} - \mathbf{r}_0)^\perp + v_\rho(\mathbf{r} - \mathbf{r}_0) \\ &+ \mathbf{v}_0 + \frac{\mathbf{r}\mathbf{r}^t - \mathbf{r}_0\mathbf{r}_0^t}{f_c^2} \mathbf{v}_0) \cdot \nabla(g_1(\mathbf{r}) - g_0(\mathbf{r})) + j\Delta_\omega g_1(\mathbf{r})|^2 d\mathbf{r}, \end{aligned} \quad (20)$$

which is solved by inverting the normal equation

$$\Re \left\{ \int \int_{\Gamma(\mathbf{r})} \begin{bmatrix} h_\theta^* \\ h_\rho^* \end{bmatrix} \begin{bmatrix} h_\theta & h_\rho \end{bmatrix} d\mathbf{r} \right\} \begin{bmatrix} \Omega_z \\ v_\rho \end{bmatrix}$$

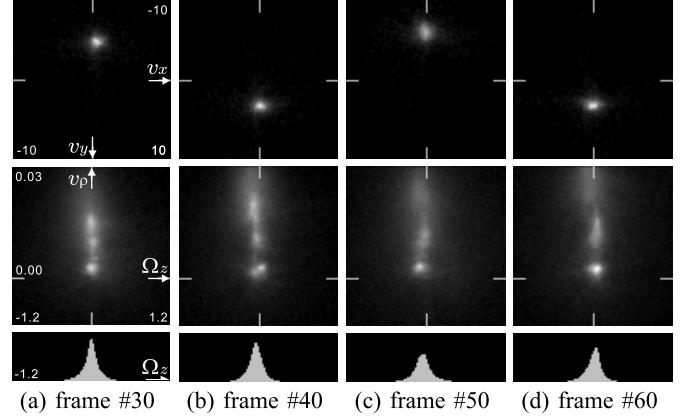


Fig. 7. Examples of histograms for the estimation of gaze parameters \mathbf{v}_0 and Ω_z . See Figs. 12(a) to (d) top row for those images. Upper row: histograms of (v_x, v_y) in the gaze rectangle (axis scale: pixel/frame), middle row: histograms of (Ω_z, v_ρ) in the peripheral image field (axis scale: deg/frame and 1/frame), and bottom row: vertical projections of the (Ω_z, v_ρ) histogram for all $v_\rho > 0$ bins (axis scale: deg/frame).

$$\begin{aligned} &= -\Re \left\{ \int \int_{\Gamma(\mathbf{r})} \begin{bmatrix} h_\theta^* \\ h_\rho^* \end{bmatrix} \left(j\Delta_\omega g_1 \right. \right. \\ &\quad \left. \left. + (\mathbf{v}_0 + \frac{\mathbf{r}\mathbf{r}^t - \mathbf{r}_0\mathbf{r}_0^t}{f_c^2} \mathbf{v}_0) \cdot \nabla(g_1 - g_0) \right) d\mathbf{r} \right\}, \end{aligned} \quad (21)$$

where

$$h_\theta \equiv (\mathbf{r} - \mathbf{r}_0)^\perp \cdot \nabla(g_1 - g_0) \quad (22)$$

$$h_\rho \equiv (\mathbf{r} - \mathbf{r}_0) \cdot \nabla(g_1 - g_0). \quad (23)$$

The results in the peripheral regions make a 2D histogram. Even for static objects, $v_\rho = T_z/Z$ varies in accordance with the distance. A marginal distribution of it for positive v_ρ is obtained and then weighted by the similarity with the previous frame. The peak is used as the estimate $\tilde{\Omega}_z^{[m]}$. By limiting v_ρ to be positive, regions of moving objects faster than the platform are excluded. A Kalman filter is used to obtain a smoothed pair of $\Omega_z^{[m]}$ and $\theta_z^{[m]}$ where θ_z and $\theta_z^{[m]}$ are the roll angles.

The top and middle rows in Fig. 7 show examples of 2D velocity histograms for estimating \mathbf{v}_0 and Ω_z , respectively. In the top-row histograms, single clear peaks moving up and down are evident. The peaks usually correspond to a stationary distant scene or the back face of a large vehicle in front. The highest peak is chosen as \mathbf{v}_0 unless otherwise specified, and the contributing region to it is extracted as the gaze area. The middle row shows the 2D histograms of Ω_z (horizontal axis) and v_ρ (vertical axis) in the peripheral image region. In it, populations on the upper side are from proximate objects and those near the center are from a distant scene. Graphs in the bottom row show the vertical projection of the middle-row histogram. Their peaks are used as Ω_z .

Fig. 8 shows an example of velocity detection and gaze-target tracking when the vehicle is passing through an intersection under repair. Images are shown in Fig. 12. In the top and bottom graphs, the traces of the gaze position x_0, y_0, θ_z and the gaze velocity v_{0x}, v_{0y}, Ω_z are shown, respectively. Gray lines near each trace show the gyro sensor outputs as the ground truth values. In the graphs, repetitive changes in pitch angle and motion owing to the waviness of the road surface are

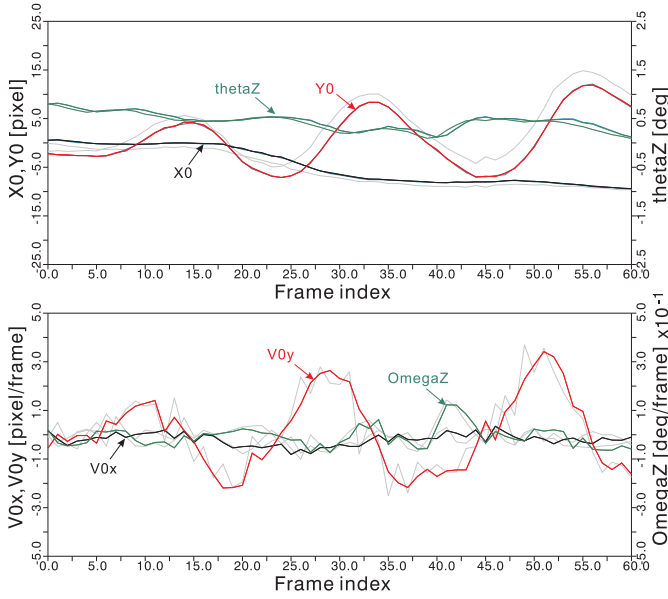


Fig. 8. Examples of traces of gaze parameters $\mathbf{r}_0 = (x_0, y_0), \theta_z$, $\mathbf{v}_0 = (v_{0x}, v_{0y})$ and Ω_z (for corresponding images, see Figs. 12(a) to (d) top row) Top graph: x_0 (black), y_0 (red) and θ_z (green); bottom graph: v_{0x} (black), v_{0y} (red) and Ω_z (green). Gray lines near the traces show the gyro sensor outputs as the ground truth values.

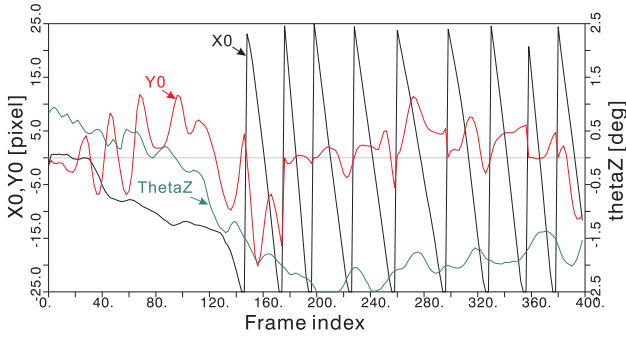


Fig. 9. Long-term trace of the gaze center position $\mathbf{r}_0 = (x_0, y_0)$ and the roll angle θ_z . The graphs of frame indices from 13 to 73 are the same as those from 0 to 60 in Fig. 8. The lines are x_0 (black), y_0 (red) (left scale) and θ_z (green) (right scale). Repeated jumps of x_0 indicate the saccadic change in the gaze target along a curving road.

clearly captured as the traces of y_0 and v_{0y} . Velocity traces are very accurate and delay-free: about 0.2 pixel/frame for v_{0x}, v_{0y} (0.0086 deg/frame as Ω_x and Ω_y) and 0.1 deg/frame for Ω_z . The increasing offset of y_0 is due to the continuing slope of the road (slanted ground plane).

The tracking operation continues at all times while keeping or inevitably changing the gaze target. Fig. 9 shows a long-term trace of (x_0, y_0) and θ_z after the scenes in Fig. 12. Repeated jumps of x_0 indicate the saccadic changes of gaze target along a curving road so that the gaze rectangle is near the image center. Slight discontinuities of y_0 at the jumps are caused by the shift of gaze center to the most textured position.

D. Estimation of GazeFlow

As the final step in each frame, the gazeFlow is estimated directly from the image data $g_0(\mathbf{r})$ and $g_1(\mathbf{r})$ using the

above determined gaze parameters \mathbf{v}_0 , \mathbf{r}_0 and Ω_z . The Horn-Schunck-type global optimization method [10] is used for the calculation. A particular emphasis is on the regularization of ill conditions using an a priori distribution given by a previous frame and the knowledge of stationary objects such as the ground plane and sky. Owing to the enhanced stability and continuity in the SVC, the regularization performs well along both the temporal and spatial axes.

Let the a priori gazeFlow distribution be $\tilde{\mathbf{v}}_G(\mathbf{r})$. Then, the estimation problem is stated as the minimization of the sum of the squared deviation of the gazeFlow from the default gazeFlow, the squared deviation from the local average, and the squared error of OFAE:

$$\begin{aligned} & \{\mathbf{v}_G(\mathbf{r})\} \quad (\text{gazeFlow distribution}) \\ & = \arg \min_{\{\mathbf{v}_G(\mathbf{r})\}} \int \int_{\text{whole image}} \left\{ \|\mathbf{v}_G(\mathbf{r}) - \tilde{\mathbf{v}}_G(\mathbf{r})\|^2 \right. \\ & \quad + \mu^2 \|\mathbf{v}_G(\mathbf{r}) - \langle \mathbf{v}_G(\mathbf{r}) \rangle - \langle \tilde{\mathbf{v}}_G(\mathbf{r}) \rangle\|^2 \\ & \quad + \lambda^2 \left[(\mathbf{v}_G(\mathbf{r}) + \mathbf{v}_0 + \frac{\mathbf{r}\mathbf{r}^t - \mathbf{r}_0\mathbf{r}_0^t}{f_c^2} \mathbf{v}_0 + (\mathbf{r} - \mathbf{r}_0)^\perp \Omega_z) \right. \\ & \quad \left. \cdot \nabla(g_1(\mathbf{r}) - g_0(\mathbf{r})) + j \Delta_\omega g_1(\mathbf{r}) \right]^2 \left. \right\} d\mathbf{r}, \quad (24) \end{aligned}$$

where $\{\mathbf{v}_G(\mathbf{r})\}$ indicates the set of gazeFlows in the whole image, $\langle \cdot \rangle$ indicates a spatial integral for smoothing, and μ^2 and λ^2 are the regularization parameters. Let us express $\tilde{\mathbf{v}}_G(\mathbf{r}) = \tilde{\mathbf{v}} = (\tilde{v}_x, \tilde{v}_y)$ and $\mathbf{v}_G(\mathbf{r}) = \tilde{\mathbf{v}} + \delta\mathbf{v} = (\tilde{v}_x + \delta v_x, \tilde{v}_y + \delta v_y)$. Let $(h_x, h_y) \equiv \nabla(g_1(\mathbf{r}) - g_0(\mathbf{r}))$, $h_\theta \equiv (\mathbf{r} - \mathbf{r}_0)^\perp \cdot \nabla(g_1(\mathbf{r}) - g_0(\mathbf{r}))$, and $\mathbf{v}_0 + (\mathbf{r}\mathbf{r}^t - \mathbf{r}_0\mathbf{r}_0^t)\mathbf{v}_0/f_c^2 \equiv (v'_{0x}, v'_{0y})$ for brevity. Then, the integral is expressed as

$$\begin{aligned} J = \int \int_{\text{whole image}} & (\delta v_x^2 + \delta v_y^2 + \mu^2 (\delta v_x - \langle \delta v_x \rangle)^2 \\ & + \mu^2 (\delta v_y - \langle \delta v_y \rangle)^2 + \lambda^2 |h_x(\tilde{v}_x + \delta v_x + v'_{0x}) \\ & + h_y(\tilde{v}_y + \delta v_y + v'_{0y}) + h_\theta \Omega_z + j \Delta_\omega g_1|^2) d\mathbf{r}. \quad (25) \end{aligned}$$

By differentiating J with δv_x and δv_y for all \mathbf{r} and equating to zero, we obtain the iterative scheme

$$\begin{aligned} \delta v_x^{k+1} = & \frac{\mu^2}{1 + \mu^2} \langle \delta v_x^k \rangle - \frac{\lambda^2}{1 + \mu^2} \Re \{ h_x^* (h_x (\tilde{v}_x + \delta v_x^k + v'_{0x}) \\ & + h_y (\tilde{v}_y + \delta v_y^k + v'_{0y}) + h_\theta \Omega_z + j \Delta_\omega g_1) \} \quad (26) \end{aligned}$$

$$\begin{aligned} \delta v_y^{k+1} = & \frac{\mu^2}{1 + \mu^2} \langle \delta v_y^k \rangle - \frac{\lambda^2}{1 + \mu^2} \Re \{ h_y^* (h_x (\tilde{v}_x + \delta v_x^k + v'_{0x}) \\ & + h_y (\tilde{v}_y + \delta v_y^k + v'_{0y}) + h_\theta \Omega_z + j \Delta_\omega g_1) \}, \quad (27) \end{aligned}$$

where the superscripts k and $k+1$ are the iteration counts. For the regularization parameters, our choice in the experiments was $\mu^2 \sim 20$ and

$$\frac{\lambda^2}{1 + \mu^2} \sim \frac{1}{(|h_x|^2 + |h_y|^2 + \varepsilon_1^2) \sqrt{(\delta v_x^k)^2 + (\delta v_y^k)^2 + \varepsilon_2^2}}, \quad (28)$$

where ε_1^2 and ε_2^2 are small positives.

The computation time for each frame (3PCIS @30fps) from the image input, the gaze and imaging operations, and to the display of results was about 0.62s using a Core™ i7-4790 CPU @3.6GHz (Visual C++ 14.1 codes, single thread). A suitable GPU will enable the real-time implementation.

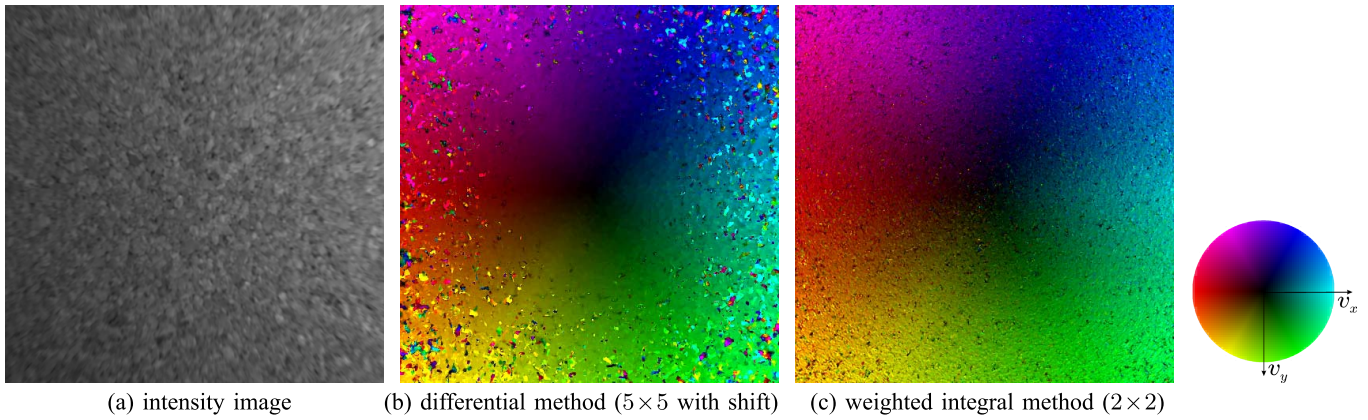


Fig. 10. Optical flow detection of an approaching wall. All images were taken by a 3PCIS [26]. (a) intensity image (first frame for the DM), (b) OF result of the DM (two frames, 5×5 pixel least squares, cascaded estimates using low resolution images and pixel-shifted high-resolution images), (c) OF result of the WIM (single frame, 2×2 pixel least squares using Eq. (15)). Velocity vs color correspondence is indicated by the color scale to the right of (c).

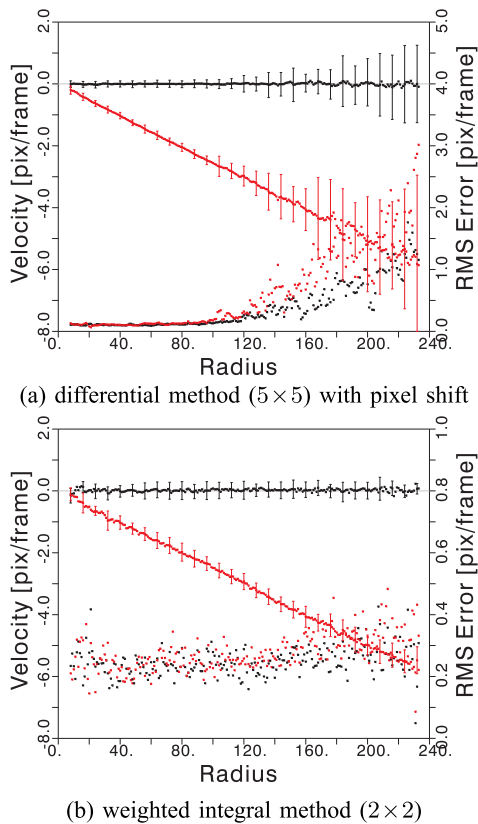


Fig. 11. Tangential (black) and radial (red) velocity components (left-axis scale) and rms estimation errors (right-axis scale) of the OF results in Figs. 10(b) and (c). The horizontal axis is the distance from the focus of expansion. Vertical bars indicate the rms error. Note that the right-axis scale (rms error) in (a) is magnified 5-fold that in (b).

VI. EXPERIMENTAL EVALUATION

A. Performance Evaluation of OF Detection by 3PCIS

The proposed method is based on the extended sensing capability of the 3PCIS and the exact algebraic solution based on the WIM. We first confirm the performance of this device and algorithms with a simple ego-motion setup. Fig. 10 shows OF results of a flat board (cork) approaching

the camera. It simulates all directions and wide magnitudes of OF from the vehicle (see [26] for the results obtained using a rotating object). The intensity image (a) captured at 30.3 frame/s shows $g_0(\mathbf{r})$. It is also used as the first frame of a simple dual-resolution DM [8]. Spatial gradients (2-level smoothing for the DM) were computed in the frequency domain for accuracy reasons, while 5×5 and 2×2 pixels local least-squares methods were applied for the DM and WIM, respectively. In the DM, pixel shift was introduced to reduce the relative displacement using the first velocity estimate. However, the DM result Fig. 10(b) shows large disorders except for the central region with low velocities (~ 3 pixels/frame). For an instantaneous OF result, velocity errors such as spots are difficult to distinguish from small moving objects. In contrast, the stability is maintained even at the peripherals (~ 6 pixels/frame) in the WIM result Fig. 10(c). Moreover, the temporal and spatial resolutions are 2 and 2.5 times larger than those in Fig. 10(b). Small decreases in velocity in Fig. 10(c) are caused by the aperture problem: the decrease in $\nabla(g_0(\mathbf{r}) - g_1(\mathbf{r}))$ particularly under high velocity conditions [22], [23]. Figs. 11(a) and (b) show the velocity dependences of the accuracy of the DM and WIM. In both results, the tangential and radial velocity components, together with the rms errors are plotted. The DM has an unavoidable velocity limit even when smoothing the images. In contrast, the accuracy of the WIM is maintained over a wide velocity range. Even with a small estimation area (2×2 pixels), the rms error at a low velocity is about 0.25 pixels/frame and the relative error at a high velocity is about 4%. This shows that high-resolution, wide-range velocity can be obtained directly and more reliably by using the 3PCIS combined with the WIM.

B. Gaze-flow Detection From a Moving Vehicle

Figs. 12(a) to (d) show conventional egocentric OF and gaze-flow. Both were obtained by the 3PCIS with the WIM. The scenes were acquired when the vehicle is passing through an intersection under repair. A few workers are standing on the road near construction equipments. In the second row, egocentric OFs are shown. The velocity and directions

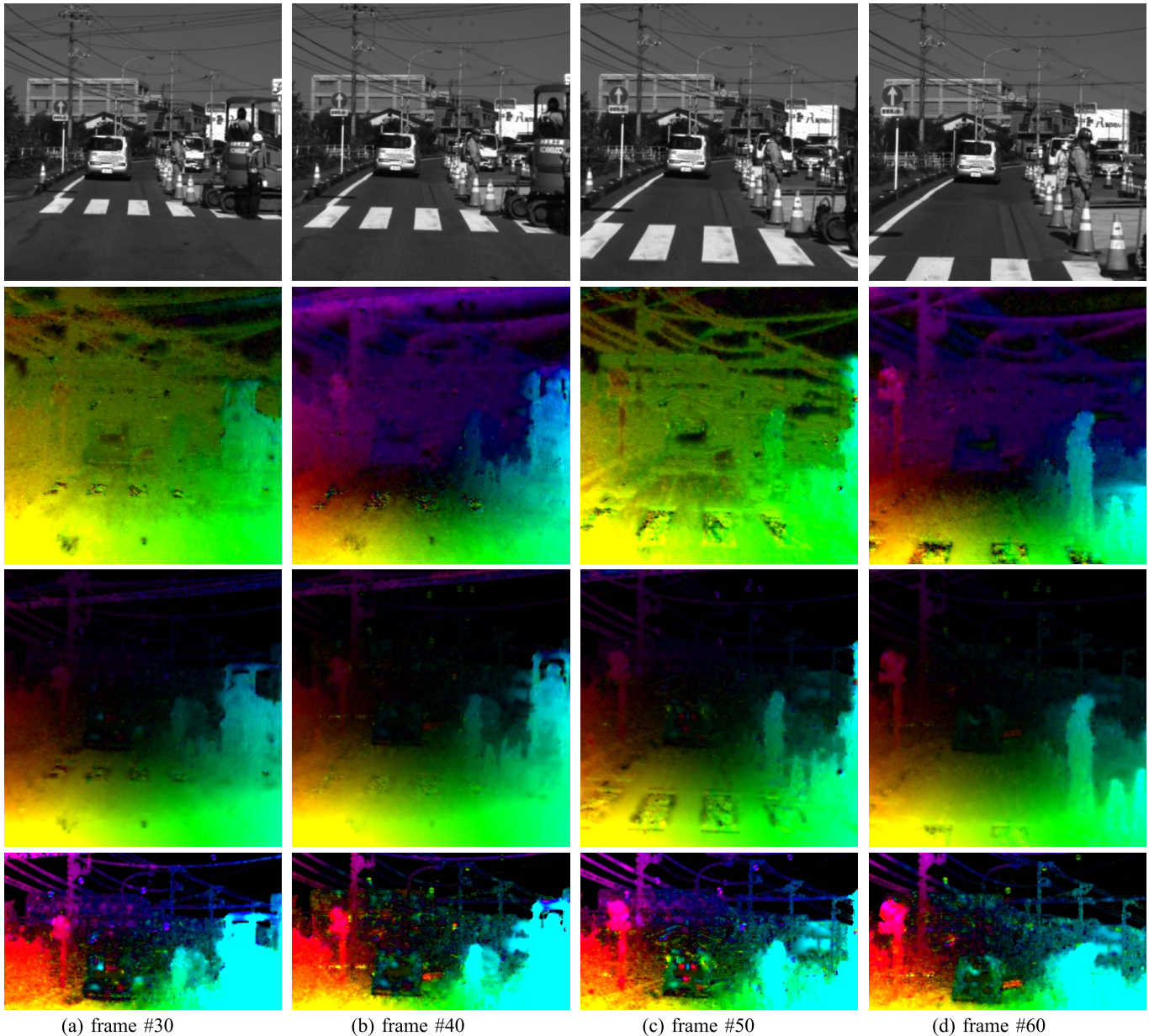


Fig. 12. Egocentric OF and gaze-flow (both were obtained by 3PCIS with WIM). The test vehicle with 3PCIS is passing through an intersection under repair. First row: intensity images; second row: camera-centered OFs; third row: gaze-flows (maximum brightness: 5.0 pixel/frame), fourth row: magnified brightness of the third row (ditto: 1.0 pixel/frame). The egocentric OF changes severely and rapidly owing to the ego-motion. In contrast, most of the ego-motion components are removed, and the distance-related flow field caused by the vehicle advance and its details become evident.

indicated by color change severely and rapidly owing to the pitching and rolling of the vehicle. In contrast, the gaze-flows shown in the third row are very stable and present a mostly gradual expansion of a stationary distribution. This is an inverse-distance-related flow field caused by the locomotion of the vehicle. In it, the workers and construction equipment can be clearly recognized as proximate objects in the advancing direction. The velocities on the road are more stabilized in the gaze-flow because the regularization effects in the iteration are enhanced by the temporal continuity of the velocity field in this area. The regularization also clarifies the small OF in distant scenes as shown in the fourth row (maximum brightness: 1.0 pixel/frame). Increasing rightward velocities of power poles are from 0.2 to 0.8 pixel/frame.

Changing colors due to specular reflections are seen on the vehicle in front. Other bright spots are mostly due to sensor dusts.

The accuracy of OF and gaze-flow was mostly maintained over an entire velocity range of this experimental condition. Even without an interframe regularization, the rms error at a low velocity is about 0.25 pixel/frame and the relative error at a high velocity is about 4%.

C. Anomaly Detection From Moving Vehicle

Figs. 13(a) to (l) show several scenes and results including moving objects. (a), (b) and (c) are images on a branch of a town road with much traffic. In gaze-flow (b), a bicycle

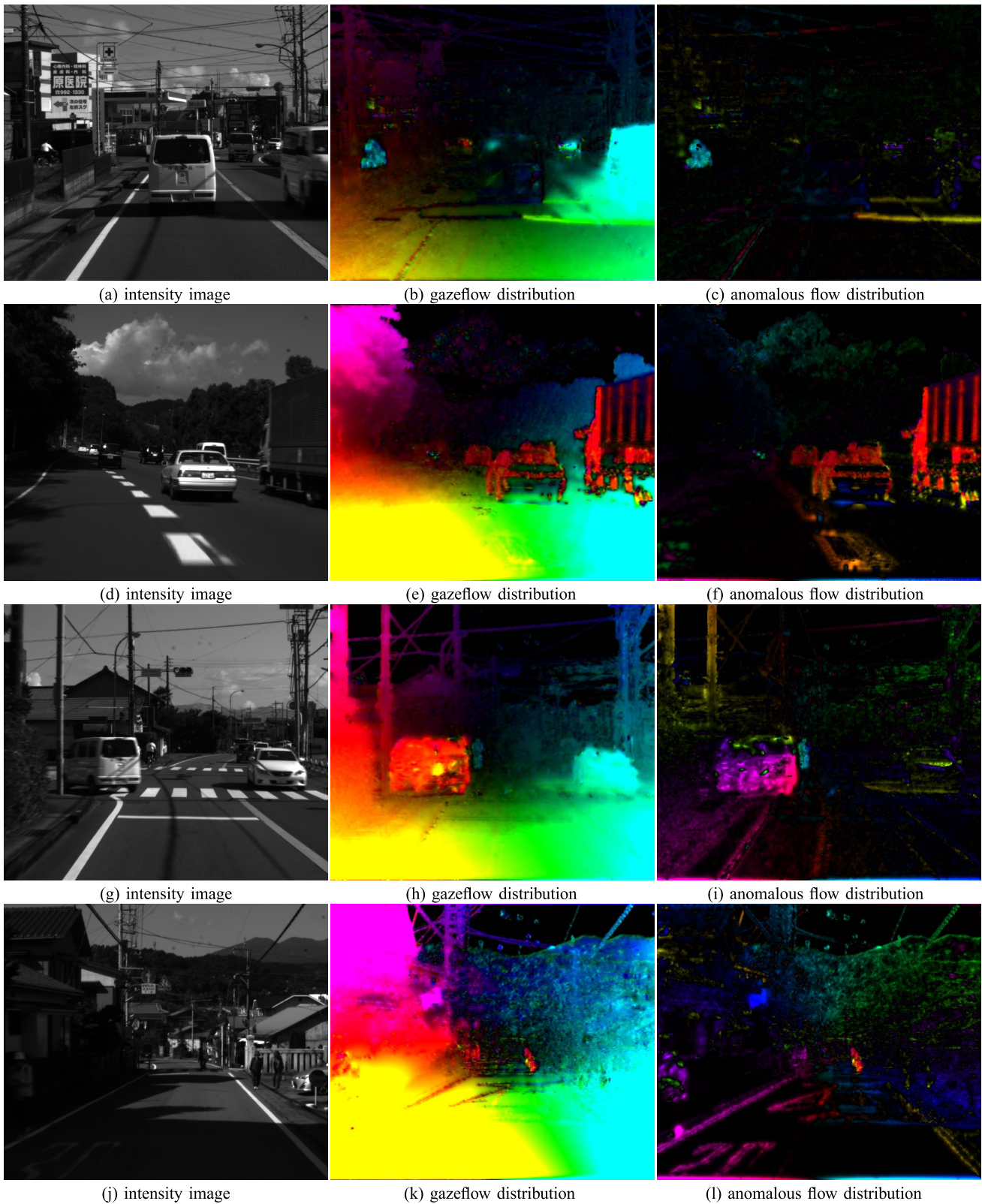


Fig. 13. Examples of gaze flow and anomalous flow distributions of 3D scenes including moving objects. (a),(b),(c) A town road with intense traffic. Several motions of on-coming cars and a cycling student approaching the road from left are detected as velocities distinct from the environment. As shown in (c), the rightward motion of the bicycle cannot be attributed to a stationary object. (d),(e),(f) A merging lane to a highway. Passing vehicles faster than the test vehicle are captured as distinctive and contradictive motions to the stationary environment. (g),(h),(i) A left-turning vehicle and a bicycle hidden by it. The leftward velocity of the vehicle and its tangential component are captured in (h) and (i). The rightward velocity of the bicycle generates an anomalous flow. (j),(k),(l) Pedestrians walking across or along a narrow town road. The leftward velocity of a crossing pedestrian (red) captured in (k) is extracted as an anomalous flow in (l). Some tangential components of body motions of a walking pedestrian (left side) are also captured as anomalous flows.

TABLE II

FEATURES AND ROLES OF GAZEFLOW IN ON-VEHICLE IMAGE ANALYSIS FOR SAFETY AND OBSTACLE AVOIDANCE (DOF: DEGREE OF FREEDOM, OV: OWN VEHICLE, EW: EARLY WARNING). REGARDING NOTES 1) TO 5), SEE THE DISCUSSION AT THE END OF SECTION VI-C

applied conditions and objects	responses of egocentric optical flow	responses of gaze-flow
general 3D environment	ego-motion (6DOF) + object motion	simple ego-motion (~ 1 DOF) + object motion
stationary surface	ego-motion (6DOF)	diverging flow proportional to inverse distance
moving object	ego-motion (6DOF) + object motion	diverging flow proportional to inverse distance + object motion tangential flow due to object motion (suspicious ¹⁾)
front vehicle (equal speed to OV)	same or different flow from surrounds	zero or lateral/turning motion (suspicious ²⁾)
front vehicle (faster than OV)	different flow from surroundings	negative diverging flow (anomalous) + lateral/turning motion (suspicious ²⁾)
front vehicle (slower than OV)	different flow from surroundings	larger diverging flow than surround (suspicious) + lateral/turning motion (suspicious ²⁾) nearer object under static assumption (safe side as EW)
parallel vehicle (faster than OV)	different flow from surroundings	negative diverging flow (anomalous ³⁾)
parallel vehicle (slower than OV)	different flow from surroundings	positive diverging flow smaller than surroundings (suspicious) distant object under static assumption (danger side as EW)
on-coming vehicle	different flow from surroundings	positive diverging flow larger than surroundings (suspicious ⁴⁾) nearer object under static assumption (safe side as EW)
laterally approaching object	different flow from surroundings	negative diverging flow (anomalous ⁵⁾)
laterally leaving object	different flow from surroundings	positive diverging flow larger than surroundings (suspicious) nearer object under static assumption (safe side as EW)



Fig. 14. Induced transition of the gaze target to a vehicle crossing in front. (a) and (b) gaze areas (red) in gaze rectangles (cyan) before and after the transition, (c) and (d) gaze-flows for (a) and (b). In (a), the gaze target was on the left and through the window of the vehicle, and the gaze-flow (c) captures the leftward motion of the crossing vehicle. In (d), relative motions in the vehicle (wheel rotation and surrounding reflection) become more evident.

approaching from the left is captured as an area of rightward velocity. Also, the rightward velocity of on-coming cars and a low leftward velocity of a right-turning car in the distant intersection are very distinctive from the surroundings. They are suspicious because of their shorter times to collision [62] than the surroundings. The anomalous flow (c) displays the tangential component of gaze-flow or the gaze-flow as it is where the diverging component is negative. Only the approaching bicycle is visible in it. To extract the on-coming and

right-turning cars as moving objects, their distance relations with the surroundings must be elucidated. Figs. 13(d) to (f) represent a scene and the flow results for a merging lane of a highway. In the gaze-flow (e), several passing vehicles that are faster than the test vehicle show opposite velocities to the stationary environment, and only the passing cars are visible in the anomalous flow (f). Figs. 13(g) to (i) are for a scene and the results of a forward vehicle turning left, whereby bicycles (rightward velocity ~ 1.2 pixel/frame) hidden by it are revealed. The leftward velocities of the vehicle higher than the surroundings are captured in the gaze-flow (h). Also, their tangential components on its lower side are captured as the anomalous flow (i). One of the bicycles is moving slightly rightward along the road. Therefore, it is nonzero in the gaze-flow and is also detected as an anomalous flow. In Figs. 13(j), (k) and (l), a pedestrian crossing the road in front is captured in the gaze-flow and anomalous flow. The leftward velocity on the right side of the road is improbable as a diverging flow from a moving vehicle. Tangential components of the walking motion of a pedestrian on the left side are also captured as anomalous flows, whereas two slow walkers on the right side of the road are not. The other nonzero components in (k) are caused by the insufficient cancellation of ego-motion OFs and/or aperture problems in textureless regions or oriented patterns.

Fig. 14 shows a passive transition of the gaze target when a vehicle crossed the path in front. The gaze changed from a road side to a side face of the moving vehicle, which is dominant in the gaze rectangle. This causes a change of gaze-flows since they are relative to the target motion. In Fig. 14(c), the gaze-flow captures clearly the leftward motion of a crossing vehicle. Also, ego-motion components are suppressed mostly so that the sensitivity to other moving objects is maintained. In contrast, Fig. 14(d) captures detailed motions of the vehicle parts including their directions, e.g. a rotating wheel of the vehicle. However, a uniform bias is present in the OFs of environment and the removal of ego-rotation components is

degraded. The suppression or utilization of this type of gaze transition can be a subject of future studies.

TABLE II summarizes the features of gazeflow for acquiring the 3D environment and moving objects in it. The advantages of gazeflow over the conventional egocentric OF are evident in the rapid finding of moving objects as anomalous or suspicious events. As early warnings of danger, these events can be used to trigger the suitable high-level image analysis. For the early warning conditions from the notes 1) to 5) in TABLE II, examples of gazeflow and anomalous flow responses are as follows: 1) tangential flow due to object motion: Fig. 13(l) (walker's leg on the left); 2) the lateral or turning motion of the forward vehicle: Figs. 13(h) and (i); 3) negative diverging flow by a faster parallel vehicle: Figs. 13(e) and (f); 4) positive diverging flow larger than the surroundings: Figs. 13(b) and (h); 5) laterally approaching or crossing objects: Figs. 13(b), (c) and (l).

VII. SUMMARY

An on-vehicle OF detection scheme with an extended spatiotemporal resolution of the flow field and stability against the rotational fluctuation of the platform was proposed based on the exact algorithm and the solid-state 3PCIS with Fourier-coefficient imaging capability. A gaze operation was introduced so that the OF field relative to the object-centric coordinate, called the gazeflow, is stabilized and the detailed motion of small objects becomes detectable. The overall performance including the anomalous object extraction was examined using real 3PCIS data sequences acquired by a vehicle moving on busy town roads and a highway. The proposed method is expected to be helpful for enhancing the performance of succeeding high-level operations and for realizing advanced safety vehicles with keen senses.

REFERENCES

[1] C. Urmson *et al.*, "Autonomous driving in urban environments: Boss and the urban challenge," *J. Field Robot.*, vol. 25, no. 1, pp. 425–466, Jun. 2009.

[2] M. S. Darms, P. E. Rybski, C. Baker, and C. Urmson, "Obstacle detection and tracking for the urban challenge," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 3, pp. 475–485, Sep. 2009.

[3] R. Roriz, J. Cabral, and T. Gomes, "Automotive LiDAR technology: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 6282–6297, Jul. 2021, doi: [10.1109/TITS.2021.3086804](https://doi.org/10.1109/TITS.2021.3086804).

[4] T. Kindo and B. Okumura, "Theory of collision avoidance capability in automated driving technologies," *SAE Int. J. Connected Automated Vehicles*, vol. 1, no. 2, pp. 67–80, Oct. 2018.

[5] S. Shalev-Shwartz, S. Shammah, and A. Shashua, "On a formal model of safe and scalable self-driving cars," 2017, *arXiv:1708.06374*.

[6] D. Geronimo, A. M. Lopez, A. D. Sappa, and T. Graf, "Survey of pedestrian detection for advanced driver assistance systems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 7, pp. 1239–1258, Jul. 2010.

[7] D. Sun, S. Roth, and M. J. Black, "A quantitative analysis of current practices in optical flow estimation and the principles behind them," *Int. J. Comput. Vis.*, vol. 106, no. 2, pp. 115–137, 2014.

[8] D. F. P. Bouthemy and C. Kervrann, "Optical flow modeling and computation: A survey," *Comput. Vis. Image Understand.*, vol. 134, pp. 1–21, May 2015.

[9] C. Cafforio and F. Rocca, "Methods for measuring small displacements of television images," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 5, pp. 573–579, Sep. 1976.

[10] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artif. Intell.*, vol. 17, nos. 1–3, pp. 185–203, Aug. 1981.

[11] S. Negahdaripour and B. K. P. Horn, "A direct method for locating the focus of expansion," *Comput. Vis., Graph., Image Process.*, vol. 46, no. 3, pp. 303–326, Jun. 1989.

[12] J. Weber and J. Malik, "Robust computation of optical flow in a multi-scale differential framework," *Int. J. Comput. Vis.*, vol. 14, no. 1, pp. 67–81, Jan. 1995.

[13] S. Uras, F. Girosi, A. Verri, and V. Torre, "A computational approach to motion perception," *Biol. Cybern.*, vol. 60, no. 2, pp. 79–87, 1988.

[14] L. Zappella, A. D. Bue, X. Lladó, and J. Salvi, "Joint estimation of segmentation and structure from motion," *Comput. Vis. Image Understand.*, vol. 117, no. 2, pp. 113–129, Feb. 2013.

[15] C. Vogel, K. Schindler, and S. Roth, "3D scene flow estimation with a piecewise rigid scene model," *Int. J. Comput. Vis.*, vol. 115, no. 1, pp. 1–28, 2015.

[16] S. Ando and A. Kimachi, "Correlation image sensor: Two-dimensional matched detection of amplitude modulated light," *IEEE Trans. Electron Devices*, vol. 50, no. 10, pp. 2059–2066, Oct. 2003.

[17] S. Ando and T. Nara, "An exact direct method of sinusoidal parameter estimation derived from finite Fourier integral of differential equation," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3317–3329, Sep. 2009.

[18] T. Nara and S. Ando, "Direct localization of poles of meromorphic function from measurements on incomplete boundary," *Inverse Problems*, vol. 26, Jan. 2010, Art. no. 015011.

[19] S. Ando, T. Nara, and T. Levy, "Partial differential equation-based localization of a monopole source from a circular array," *J. Acoust. Soc. Amer.*, vol. 134, no. 4, pp. 2799–2813, 2013.

[20] S. Ando, T. Nara, and T. Kurihara, "Spatial filtering velocimetry revisited: Exact short-time detecting schemes from arbitrarily small-size reticles," *Meas. Sci. Technol.*, vol. 25, no. 8, Aug. 2014, Art. no. 085001.

[21] S. Ando, "Frequency-domain Prony method for autoregressive model identification and sinusoidal parameter estimation," *IEEE Trans. Signal Process.*, vol. 68, pp. 3461–3470, 2020.

[22] S. Ando, W. Dabi, and P. Masurel, "Optical flow detection via complex-sinusoidally modulated imaging and its realization with correlation image sensor," *Trans. IPSJ*, vol. 49, no. 6, pp. 13–21, 2008.

[23] S. Ando, T. Kurihara, and D. Wei, "Exact algebraic method of optical flow detection via modulated integral imaging," in *Proc. Int. Conf. Comput. Vis. Theory Appl. (VISAPP)*, Lisboa, Portugal, 2009, pp. 480–487.

[24] S. Han, T. Sawada, T. Iwahori, S. Kawahito, and S. Ando, "A three-phase time-correlation image sensor using pinned photodiode active pixels," *Proc. SPIE*, vol. 7536, pp. 244–250, Jan. 2010.

[25] M. Habibi, "Analysis, enhancement, and sensitivity improvement of the correlation image sensor," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 3, pp. 708–718, Mar. 2012.

[26] S. Ando, M. Nagase, T. Watanabe, T. Kosugi, and T. Iida, "Correlation image sensor for algebraic solution of optical flow," *IEEE J. Electron Devices Soc.*, vol. 9, pp. 170–179, 2021.

[27] H. Longuet-Higgins and K. Prazdny, "The interpretation of the moving retinal image," *Proc. Roy. Soc. London B*, vol. 208, pp. 385–397, Jan. 1980.

[28] B. K. P. Horn and E. J. Weldon, "Direct methods for recovering motion," *Int. J. Comput. Vis.*, vol. 2, no. 1, pp. 51–76, Jun. 1988.

[29] M. Irani, B. Rousso, and S. Peleg, "Recovery of ego-motion using region alignment," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 3, pp. 268–272, Mar. 1997.

[30] D. Scaramuzza, "1-point-RANSAC structure from motion for vehicle-mounted cameras by exploiting non-holonomic constraints," *Int. J. Comput. Vis.*, vol. 95, no. 1, pp. 74–85, 2011.

[31] O. Lanz, "Approximate Bayesian multibody tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 9, pp. 1436–1449, Sep. 2006.

[32] T. Basha, Y. Moses, and N. Kiryati, "Multi-view scene flow estimation: A view centered variational approach," *Int. J. Comput. Vis.*, vol. 101, no. 1, pp. 6–21, 2013.

[33] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: Part I," *IEEE Robot. Autom. Mag.*, vol. 13, no. 2, pp. 99–110, Jun. 2006.

[34] G. P. Stein, O. Mano, and A. Shashua, "A robust method for computing vehicle ego-motion," in *Proc. IEEE Intell. Vehicles Symp.*, Oct. 2011, pp. 362–368.

[35] A. Geiger, M. Lauer, C. Wojek, C. Stiller, and R. Urtasun, "3D traffic scene understanding from movable platforms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 5, pp. 1012–1025, May 2014.

[36] H. Gonzalez, S. Rodriguez, and A. Elouardi, "Track-before-detect framework-based vehicle monocular vision sensors," *Sensors*, vol. 19, no. 3, pp. 560–584, 2019.

- [37] A. Ess, B. Leibe, K. Schindler, and L. Van Gool, "Robust multiperson tracking from a mobile platform," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 10, pp. 1831–1846, Oct. 2009.
- [38] R. Sabzevari and D. Scaramuzza, "Multi-body motion estimation from monocular vehicle-mounted cameras," *IEEE Trans. Robot.*, vol. 32, no. 3, pp. 638–651, Jun. 2016.
- [39] W. Ci and Y. Huang, "A robust method for ego-motion estimation in urban environment using stereo camera," *Sensors*, vol. 16, no. 10, p. 1704, Oct. 2016.
- [40] T. Chen and S. Lu, "Object-level motion detection from moving cameras," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 11, pp. 2333–2343, Nov. 2017.
- [41] U. Franke and S. Heinrich, "Fast obstacle detection for urban traffic situations," *IEEE Trans. Intell. Transp. Syst.*, vol. 3, no. 3, pp. 173–181, Sep. 2002.
- [42] J. Schmudderich *et al.*, "Estimating object proper motion using optical flow, kinematics, and depth information," *IEEE Trans. Syst., Man, Cybern.*, vol. 38, no. 4, pp. 1139–1151, Aug. 2008.
- [43] D. M. Gavrilu and S. Munder, "Multi-cue pedestrian detection and tracking from a moving vehicle," *Int. J. Comput. Vis.*, vol. 73, no. 1, pp. 41–59, Jun. 2007.
- [44] M. Enzweiler and D. M. Gavrilu, "Monocular pedestrian detection: Survey and experiments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 12, pp. 2179–2195, Dec. 2009.
- [45] J. Hariyono, V.-D. Hoang, and K.-H. Jo, "Moving object localization using optical flow for pedestrian detection from moving vehicle," *Sci. World J.*, vol. 2014, Jul. 2014, Art. no. 196415.
- [46] T. Gandhi and M. M. Trivedi, "Pedestrian protection systems: Issues, survey, and challenges," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 3, pp. 413–430, Sep. 2007.
- [47] R. O. Chavez-Garcia and O. Aycard, "Multiple sensor fusion and classification for moving object detection and tracking," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 2, pp. 525–534, Feb. 2016.
- [48] Y. Yuan, D. Wang, and Q. Wang, "Anomaly detection in traffic scenes via spatial-aware motion reconstruction," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 5, pp. 1198–1209, May 2017.
- [49] V. D. Nguyen, H. Van Nguyen, D. T. Tran, S. J. Lee, and J. W. Jeon, "Learning framework for robust obstacle detection, recognition, and tracking," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 6, pp. 1633–1646, Jun. 2017.
- [50] N. Morales, J. Toledo, L. Acosta, and J. Sanchez-Medina, "A combined voxel and particle filter-based approach for fast obstacle detection and tracking in automotive applications," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 7, pp. 1824–1834, Jul. 2017.
- [51] S. Nedeveschi, S. Bota, and C. Tomiuc, "Stereo-based pedestrian detection for collision-avoidance applications," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 3, pp. 380–391, Sep. 2009.
- [52] M.-T. Yang and J.-Y. Zheng, "On-road collision warning based on multiple FOE segmentation using a dashboard camera," *IEEE Trans. Veh. Technol.*, vol. 64, no. 11, pp. 4974–4984, Nov. 2015.
- [53] W. Song, Y. Yang, M. Fu, F. Qiu, and M. Wang, "Real-time obstacles detection and status classification for collision warning in a vehicle active safety system," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 3, pp. 758–773, Mar. 2018.
- [54] M. Irani, B. Rousso, and S. Peleg, "Recovery of ego-motion using image stabilization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Aug. 1994, pp. 1–7.
- [55] D. Raviv and M. Herman, "A unified approach to camera fixation and vision-based road following," *IEEE Trans. Syst., Man, Cybern.*, vol. 24, no. 8, pp. 1125–1141, Aug. 1994.
- [56] F. Raudies, E. Mingolla, and H. Neumann, "Active gaze control improves optic flow-based segmentation and steering," *PLoS ONE*, vol. 7, no. 6, pp. 1–19, 2012.
- [57] G. Sandini, F. Panerai, and F. A. Miles, "The role of inertial and visual mechanisms in the stabilization of gaze in natural and artificial systems," in *Motion Vision—Computational Neural and Ecological Constraints*, J. M. Zanker and J. Zeil, Eds. Berlin, Germany: Springer-Verlag, 2001.
- [58] D. Murray and A. Basu, "Motion tracking with an active camera," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 5, pp. 449–459, May 1994.
- [59] J. H. Elder, S. J. D. Prince, Y. Hou, M. Sizintsev, and E. Oleviskiy, "Pre-attentive and attentive detection of humans in wide-field scenes," *Int. J. Comput. Vis.*, vol. 72, no. 1, pp. 47–66, Apr. 2007.
- [60] E. Mueggler, H. Rebecq, G. Gallego, T. Delbruck, and D. Scaramuzza, "The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and SLAM," *Int. J. Robot. Res.*, vol. 36, no. 2, pp. 142–149, Feb. 2017.
- [61] H. D. Escobar-Alvarez, M. Ohradzansky, J. Keshavan, B. N. Ranganathan, and J. S. Humbert, "Bioinspired approaches for autonomous small-object detection and avoidance," *IEEE Trans. Robot.*, vol. 35, no. 5, pp. 1230–1232, Oct. 2019.
- [62] R. C. Nelson and J. Aloimonos, "Obstacle avoidance using flow field divergence," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 10, pp. 1102–1106, Oct. 1989.



Shigeru Ando (Member, IEEE) received the B.E., M.E., and Dr. (Engr.) degrees in mathematical engineering and information physics from The University of Tokyo in 1974, 1976, and 1979, respectively. He joined the Faculty of Engineering, The University of Tokyo, in 1979. He has been working as an Associate Professor since 1987 and a Professor from 1995 to 2016. He is currently a Professor Emeritus at the Department of Mathematical Engineering and Information Physics and the Department of Information Physics and Computing, The University of Tokyo. His research interests are in image and signal processing, inverse problems, optical and acoustic sensing, bio-inspired visual, auditory, and tactile sensor devices and algorithms.



Toshiki Kindo (Member, IEEE) received the Dr.Sc. degree in physics from Tohoku University in 1987. He joined the Japan Atomic Research Institute (1987–1989), Panasonic Corporation (1989–2003), Japan Science and Technology Agency as a Concurrent Researcher of PRESTO (1995–1998), Toyota Motor Corporation (2004–2019), and the Kanazawa Institute of Technology as a Professor. His research interests are in artificial intelligence, including neural networks, information theory, computer vision, and autonomous driving.