Unconventional Visual Sensors for Autonomous Vehicles

You Li¹, Julien Moreau², Javier Ibanez-Guzman¹

Abstract—Autonomous vehicles rely on perception systems to understand their surroundings for further navigation missions. Cameras are essential for perception systems due to the advantages of object detection and recognition provided by modern computer vision algorithms, comparing to other sensors, such as LiDARs and radars. However, limited by its inherent imaging principle, a standard RGB camera may perform poorly in a variety of adverse scenarios, including but not limited to: low illumination, high contrast, bad weather such as fog/rain/snow, etc. Meanwhile, estimating the 3D information from the 2D image detection is generally more difficult when compared to LiDARs or radars. Several new sensing technologies have emerged in recent years to address the limitations of conventional RGB cameras. In this paper, we review the principles of four novel image sensors: infrared cameras, range-gated cameras, polarization cameras, and event cameras. Their comparative advantages, existing or potential applications, and corresponding data processing algorithms are all presented in a systematic manner. We expect that this study will assist practitioners in the autonomous driving society with new perspectives and insights.

I. INTRODUCTION

Since the dawn of the automotive industry, building a self-driving car has always been a beautiful dream. Owing to the rapid development of sensors, processors and data processing algorithms, developing autonomous vehicles quickly became one of the hottest topics for research and industry, particularly as promoted by the DARPA Grand Challenges from 2004 [1]. From ADAS (advanced driver assistance systems) to fully autonomous driving (AD), SAE International formally classified the driving automation as 6 levels [2]. Among those levels, level 2 and 3 are semi-autonomous driving, i.e. the driver is still inside the loop of vehicle control. Level 4 and 5 allow for fully autonomous driving in restricted areas and anywhere, respectively.

Perception sensors, like human eyes, are critical in scanning and digitizing environments in all levels of autonomous vehicles. The common perception sensors are visual sensors like the cameras, and depth sensors such as LiDARs [3], microwave and ultrasonic radars [4]. By using a focal plane array (FPA), a typical camera passively assesses the intensities of ambient light at certain wavelengths within its optical field-of-view (FOV). Such information is saved as an image, with ambient light intensities sampled as millions of pixel values. A normal camera operates within the visible spectrum that each pixel value is indeed a vector of visible light intensities (e.g. red, green, and blue). In this paper, we refer to the camera as a monocular RGB camera by default. LiDARs and radars are sparse active range sensors that measure the distance along with the directions of the transmitted lasers or microwaves. A LiDAR usually has higher accuracy and angular resolution than a radar, whereas a microwave radar can measure velocity using the Doppler effect. In general, cameras mimic human vision and provide rich and dense contextual information. By using range measurements, LiDARs and radars are more accurate than cameras at modeling the 3D world.

The data streams generated by perception sensors are then processed within a perception system to provide useful information for further vehicle navigation. A perception system usually outputs two layers of information: 1) semantic layer and 2) physical layer. The semantic layer detects and classifies the objects of interest (e.g. pedestrians, vehicles, lane markings, traffic lights, etc), while the physical layer gives them their 3D positions, velocities, and sizes. In general, cameras are superior in the semantic layer, while LiDARs/radars are more reliable in locating the detected objects. Accelerated by the breakthrough of deep neural networks (DNN), RGB camera included perception systems have rapidly commercialized and integrated into mass-produced cars in various driving automation levels. [5] and [6] summarize current perception systems and envisage the evolution in the future.

Despite huge successes, the limitations of RGB cameras in challenging situations have been recognized seriously. Low illumination or other adverse conditions, e.g. fog or rain, can degrade the performance. The glares generated by oncoming headlamps or mirror-like reflections could blind an image. Such image defects would lead to missed detections or unknown behaviors for a perception system, posing safety concerns. To improve the safety and robustness of ADAS/AD, several emerging imaging technologies, e.g. the infrared (IR) cameras, dynamic vision sensors (event cameras), polarization cameras, gated cameras, etc, start to get spotlights. Addressing one or more weaknesses of a conventional RGB camera, those novel image sensors bring extra benefits to complement the common cameras for a better perception system.

To understand the mechanisms of those unconventional cameras, as well as their benefits, potential applications and algorithms, this paper systematically reviews four types of camera: the infrared camera, the polarization camera, the range-gated camera and the event camera. This paper is organized as follows: The RGB cameras are reviewed in Sec. II. Then, the infrared cameras, range-gated cameras, polarization cameras, and event cameras are analyzed in Sec. III, Sec. IV, Sec. V and Sec. VI respectively.
II. Principle of Conventional RGB Camera

To understand the limitations of current RGB cameras, we briefly introduce their mechanisms and then analyze the constraints from their imaging principles.

A. The Light

Light is a type of electromagnetic (EM) waves that is formed through the interaction between electric and magnetic fields. As shown in Fig. 1 (a), an EM wave is a transverse wave composed of oscillating magnetic and electric fields that are perpendicular to each other, and to the wave’s propagation direction as well. Any type of EM wave has three fundamental properties: amplitude, wavelength, and polarization. The wavelength \( \lambda \) of visible light \( \lambda \in [400\text{nm}, 700\text{nm}] \) is only a small portion of the EM spectrum ranging from Gamma rays \( \lambda < 1\text{nm} \) to radio waves \( \lambda > 1\text{m} \), as shown in Fig. 1 (b). A common RGB camera detects only the intensities i.e. amplitudes of the captured visible light through its lens and is unable to measure polarization information. In a typical road scene, the light is primarily issued from complex interactions between the emitted light from luminous objects (e.g. sun, streetlamp, headlamp, etc), the reflected light from illuminated objects (e.g. vehicle, pedestrian, building, etc.) and the scattered light from transmission medium (e.g. foggy air).

During the daytime, the sun is the most common source of light. However, human-perceivable sunshine is just a part of the whole solar irradiance on the ground. As shown by Fig. 1 (c), the spectrum of solar irradiance approximately contains 5% ultraviolet wavelengths, 43% visible wavelengths, and 52% infrared wavelengths (values from [8]). At night, vehicle’s headlamps and streetlamps are the primary source of light [9]. However, the lighting pattern of the car headlamps is strictly regulated for safety reasons: the maximum range of low beams can only reach around 60m [10] (as shown in Fig. 1 (d)), the high beams can reach over 150m but are not allowed to be used continuously.

The targets of interest (e.g. vehicles, pedestrians, etc) are visible in the images due to the light reflection from their surfaces. Two types of reflection contribute to the imaging results: (1) diffuse reflection and (2) specular reflection. Rough surfaces, such as asphalt roads and clothing, typically produce diffuse reflections that scatter incident light in various directions. Smooth surfaces, such as metallic material or wet road, would generate specular reflections (a mirror-like reflection) in which the reflected light is concentrated in specific directions determined by the incident angle and the surface property.

In many cases, the light transmission medium (e.g. air) is assumed to be transparent. However, in adverse conditions, such as fog, rain, snow, or smoke, the floating particles would cause light scattering that results in image blur. Light scattering can be roughly classified as Mie scattering or Rayleigh scattering based on the particle size to light wavelength ratio. Rayleigh scattering occurs when the particle size is very tiny w.r.t the light wavelength: the blue color of the sky is primarily caused by the Rayleigh scattering of solar irradiance at short wavelengths (e.g. blue at the end of the visible spectrum). For particle sizes similar or larger than a wavelength, such as the water droplet in fog, Mie scattering predominates [11].

B. From Light to Digital Images

The captured light from the various sources is focused by the lens of a camera to its focal plane, where a FPA (i.e. image sensor) is placed to generate images. An image sensor is indeed a 2D array of photosites that can convert light intensities into electrical signals, which are then converted into digits. Each photosite gives a pixel of the image. A photosite is a circuit made up of a photodetector and other electronic components. Based on the photoelectric effect of semiconduc-
tor material, Photodiodes are the most commonly used components. A photodiode [12] is a semiconductor that converts light into an electrical signal. When the incident photon energy absorbed by a photodiode exceeds the bandgap of its material, electron-hole pairs (EHPs) are generated. Then, a photocurrent $I_p$ is generated that is approximately linearly proportional to illuminance intensity. A photodiode only responds to specific wavelengths depending on the semiconductor material, which includes, but is not limited to: silicon (Si), germanium (Ge), indium gallium arsenide (InGaAs). Two important metrics represent a photodiode’s sensitivity, quantum efficiency (QE) and responsivity. QE $\eta$ represents the conversion efficiency of photons to electrons. For a specific wavelength $\lambda$, QE $\eta(\lambda)$ is defined as the percentage of photons hitting the photoreactive surface that produces EHPs:

$$\eta(\lambda) = \frac{\text{Electrons Out}}{\text{Photons Input}} = \frac{r_e}{r_p}$$ (1)

The responsivity $R$ measures the electrical output per optical input. It is defined as the ratio of photocurrent output $I_p$ (in amperes) to the optical power (in watts) $P$:

$$R(\lambda) = \frac{I_p}{P} = \frac{q}{h \nu f} \approx \eta(\lambda) \frac{\lambda}{1.24} (A/W)$$ (2)

where $q$ is electron charge, $h$ is Planck’s constant and $f$ is the frequency of the optical signal. Fig. 2 shows an example of responsivity curves for three common semiconductor materials, Si, Ge, and InGaAs. Silicon is sensitive to light in the visible and near-infrared spectrum. InGaAs photodiodes can detect wavelengths ranging from 800nm to 2600nm. Connecting a photodiode with resistors and amplifiers creates a photosite that converts the photocurrent into a voltage for further signal processing. An image sensor is created by assembling millions of photosites, together with other components into a 2D array. Currently, the majority of the image sensors are silicon-based and fabricated by CMOS process, and thus are referred to as CMOS Image Sensor (CIS).

By default, image sensors output grayscale values that represent light intensity. To enable color information, a color filter array (CFA) is placed just above the image sensor, so that each pixel is sensitive to a specific color wavelength. The Bayer filter array is the most common type of CFA, consisting of repeated $2 \times 2$ RGGB (red-green-green-blue) filter kernels because the human eye is more sensitive to green light. Only one of the three colors is recorded in the raw output from a Bayer-filter integrated image sensor. In image signal processor (ISP), a demosaicing algorithm is implemented to interpolate full color (e.g. a RGB vector) for every pixel. Other types of CFA, such as RCCC or RCCB, are designed to improve traffic light detection or to increase performance under low-light conditions [14] (with C standing for wideband clear filter, i.e. no color filtering). As shown in Fig. 2, silicon-based imagers have sensitivities extending into the near-infrared. An infrared cut-off filter (IRCF) is designed to block near-infrared wavelengths for better color quality. The entire imaging pipeline is depicted in Fig. 3 (a).

C. Limitations

Tremendous progress has been made for automotive cameras; the current flagship image sensors, e.g. OmniVision’s OX08B24C\(^1\), or Sony’s IMX324\(^2\) can capture 8MP images at 40fps, and have a 120db dynamic range. However, when confronted with complex road scenarios, several limitations remain, which can be roughly classified as (1) Image degradation in adverse conditions, and (2) Lack depth information.

As described in Sec. II-A, a camera is a passive sensor that relies on captured light through a complex interaction between external luminous objects, illuminated targets, and transmission medium. When the received light exceeds the imaging capability, the image quality degrades, affecting the perception results for ADAS/AD. For example, at night, the external illuminations may be insufficient to produce a clear image. During sunny days, specular reflections may appear on the surfaces of the vehicles or the road [15] that leads to over-saturation. Under adverse weather conditions (e.g. fog, rain, or snow), the strong scattering inside the transmission medium would reduce the image’s visibility [16]. Fig. 3 (b) demonstrates such challenging scenarios for RGB cameras.

III. INFRARED (IR) CAMERA

Conventional RGB cameras only “see” the visible spectrum, as highlighted in Fig. 1 (a). When the light wavelength exceeds 700nm, it enters the “infrared (IR)” spectrum, which is invisible for human and is often divided as follows: (1). Near-infrared (NIR): wavelength ranging from 0.7$\mu$m to 1.4$\mu$m. (2), Short-wavelength infrared (SWIR): wavelength ranging from 1.4$\mu$m to 3$\mu$m. (3), Long-wavelength infrared (LWIR): wavelength ranging from 8$\mu$m to 14$\mu$m. The Mid-wavelength infrared is too rare in automotive applications to be included in this paper. The researches and developments of IR cameras for automotive usages mainly focus on NIR, LWIR, and SWIR wavelengths [17]. NIR and SWIR are “reflected infrared” wavelengths that rely on external light sources such as the sun or other infrared illuminators. NIR and SWIR imagers work similar to RGB imagers in that they directly transform photons to electrical signals. While LWIR is usually referred to as “thermal infrared”, a typical LWIR imager converts the thermal radiation to heat, which is then converted to electrical signals. LWIR cameras can image the world solely through thermal emissions and thus do not require any external sources.

\(^1\)https://www.ovt.com/sensors/OX08B4C
\(^2\)https://www.sony-semicon.co.jp/products/common/pdf/IMX324_424.pdf
Fig. 3. (a) A typical pipeline depicting how the light in a scene is converted to image pixels in a RGB camera before being processed by perception algorithms for environment understanding. The captured light from complex light reflections and scatterings is first passed the IR cut-off filter and CFA, allowing the pixels to only respond to red, green, or blue light. Then, within a controlled integration time, electronic signals are generated and converted into digital pixel values inside an image sensor, which are then post-processed in the ISP. Perception algorithms analyze the image outputs for environmental understanding. (b) Four typical difficult scenarios for a common RGB camera (from left to right): low illumination at night, glare caused by specular reflection of wet road, high contrast leads to image over-saturation, a foggy image caused by light scattering.

Fig. 4. (a) A RGB-IR imaging system: the CFA is replaced by a RGB-IR filter array to capture RGB and infrared intensities (from [18]) (b) RGB (left) and NIR (right) images for the same scene. Vegetation in NIR spectrum are "brighter" than in RGB image (from [19]).

A. NIR camera

NIR imagery shares many properties with RGB imagery: as shown in Fig. 2 (a), a silicon-based imager can still exhibit NIR sensitivity until around 1100nm. As a result, with proper modifications, a RGB camera can be converted into a NIR camera. Because the CFA still has transmission spectra that bleed into NIR wavelengths, removing the IR cut-off filter or replacing it with a NIR bandpass filter converts a consumer-grade RGB camera to a NIR camera, as demonstrated in [20] and [21]. Fig. 4 (b) shows a RGB image and a NIR image of the same road scene. NIR dedicated pixels are developed to increase the NIR sensitivity. For instance, the Nyxel technology [22] achieves 50% QE at 940nm, and 70% QE at 850nm NIR wavelength. Although other types of materials, such as InGaAs [23] may have higher sensitivity in the NIR spectrum, silicon-based image sensors are more popular due to their lower cost.

In recent years, simultaneously capturing RGB-NIR images has become popular. To achieve that purpose, the CFA, e.g. the Bayer filter, is modified to pass NIR light for specific NIR pixels. Chen et al. [24] present a four-bandpass filter array to acquire RGB-NIR images. In Lu et al.’s work [25], a 4×4 pattern containing 15 visible/NIR filters and 1 NIR-only filter, is made. Park et al. [26] and Skorka et al. [27] further discuss the color distortion and correction problems caused by the RGB-IR filter array. On the industry side, Omnivision has commercialized RGB-NIR imaging systems [18] for automotive applications, as shown in Fig. 4 (a). A more comprehensive study on RGB-IR camera design could be found in Geelen et al. [28].

By simply adding external NIR illuminators, usually NIR
LEDs (light-emitting diodes) or VCSELs (vertical-cavity surface-emitting lasers), a passive NIR camera can be converted to an active night vision system. A NIR LED produces a very broad diffused light distribution, whereas a laser produces a narrow beam. For acquiring 2D images, LEDs are more affordable and thus more popular. While VCSELs enable 3D perception applications [29], e.g. structure light based 3D reconstruction. Two popular wavelengths are 850nm and 940nm. In the early days, 850nm NIR emitters were used because of higher sensitivity than 940nm. However, human eyes can still see a deep red glow from the 850nm emitter in dark conditions. This can be uncomfortable and/or confusing. Currently, 940nm is preferred due to its complete invisibility, and fewer interferences from the natural environment, as solar IR levels at 940nm are less than half compared to 850nm (see Fig. 1 (c)) due to atmospheric absorption.

B. SWIR camera

Covering the wavelengths ranging from 1.4µm to 3µm, the SWIR images are generated by reflected SWIR light like the NIR and RGB cameras. The longer wavelengths of the SWIR spectrum would reduce the scattering effects caused by the small particles existing in the transmission medium. In theory, the SWIR wavelengths can better penetrate fog, smoke, and other adverse weather conditions. At night, the "nightglow" (a night sky radiance emitted from the relaxation of hydroxyl molecules in the atmosphere) comprising mainly SWIR wavelengths ranging from 1.4µm to 1.8µm can provide illumination for SWIR cameras [30] as well.

Though silicon-based image sensors have excellent responsiveness from visible to NIR spectrum, the bandgap properties of silicon prevent them from having sufficient sensitivity above 1.1µm. The Indium gallium arsenide (InGaAs) has a lower bandgap, making it the preferred technology for SWIR imaging [31], as shown in Fig. 2. In comparison to other semiconductor materials used in the SWIR spectrum e.g. Ge or HgCdTe, InGaAs detectors are cost-effective and high-sensitive while being operated at room temperature [32]. However, compared to silicon-based sensors, InGaAs detectors suffer issues of the higher fabrication cost and pixel detects. As a result, some efforts have been made to exploit the potential of silicon-based imagers for SWIR imaging. As in Lv et al. [33], a deep neural network is trained to approximate the response of an InGaAs sensor and then used to turn a standard silicon-based CMOS sensor into a SWIR image sensor.

C. LWIR (Thermal) camera

As a phenomenon of converting thermal energy into electromagnetic energy, all matter with a temperature greater than absolute zero emits thermal radiation. This thermal radiation does not consist of a single wavelength, yet comprises a continuous spectrum. Suppose the radiating matter is ideal, i.e. the black-body, its thermal radiation \( B \) for wavelength \( \lambda \) is a function of temperature given by Planck’s law [35]:

\[
B(\lambda, T) = \frac{2hc^2}{\lambda^5} \frac{1}{e^{hc/\lambda k_B T} - 1}
\]

where \( \lambda, h, c, K_B \) are the wavelength, the Planck’s constant, the light speed, and the Boltzmann’s constant. Most of the radiation emitted by a human body (37°C) is mainly at the wavelength of 12µm, which is located in the LWIR spectrum. That’s the reason for using a LWIR camera for pedestrian/animal detection at night.

Photon detectors are excellent in thermal imaging because they directly convert the absorbed thermal radiation into electronic changes. However, due to the prohibitively expensive cryogenic cooling systems, their applications in ordinary scenarios are severely limited. Instead, detecting radiant heat is more popular in LWIR imaging technologies.

Without the need for cooling systems, a bolometer [36] is an instrument that measures heat radiation and converts it into certain measurable quantities. Fig. 5 (a) depicts a block diagram of a bolometer. An absorber and an attached thermometer are deposited above a read-out integrated circuit (ROIC) and substrate for the reason of heat insulation. The incident LWIR radiation heats the absorber material, which is typically measured by a thermometer via resistance changes. Historically, the Salisbury screen absorber has been used for bolometers in the LWIR spectrum [37]. The vanadium oxide (VOx) or the amorphous silicon (a-Si) are the common materials for the thermometer layer because they are compatible with standard semiconductor processing technologies, as Tissot et al. [38], Yon et al. [39], Yu et al. [40]. The thermal measures are then transferred to the ROIC for further processing.

A 2D microbolometer array [41] assembled by many tiny uncooled bolometers can capture thermal images. With a much more affordable price and compact size, the microbolometer array is particularly well-suited for mobile applications such as the automobile, especially for the night scenarios Gade and Moeslund [42]. Driven by the rapid progress of semiconductor technologies and MEMS technologies, modern microbolometer arrays can capture images at 60Hz speed with 1024×768 pixels that each pixel is fabricated in 12µm size. This paper will hereafter refer to thermal or LWIR cameras as an uncooled microbolometer array-based thermal imaging systems.

D. Applications of IR Cameras in Autonomous Vehicles

Through the introduced principles, the IR cameras can extend a perception system to deal with adverse conditions and with night while avoiding disturbing light emissions for humans. The LWIR camera holds a distinct and unique position because it operates without the need for external light sources. However, on the other hand, since NIR and SWIR cameras are "reflective infrared" like RGB cameras, they provide more context information, e.g. lane markings, tests in traffic signs, etc, than LWIR cameras. Despite some initiatives, SWIR camera applications are uncommon due to the high cost of InGaAs detectors. Here, we mainly review the applications of NIR and LWIR cameras in autonomous vehicles, and automotive night vision system (NVS) is one of the key areas utilizing NIR or LWIR cameras.

Many comparisons and discussions have taken place between the active NIR cameras and passive LWIR night vision systems as in Kallhammer [43]. Coupled with invisible NIR
transmitters (e.g. NIR LEDs or headlamps containing NIR spectrum), an active NIR camera is a cost-effective NVS. While LWIR cameras are particularly suitable for detecting hot-blooded creatures (humans, animals, etc) and other objects with heat signatures (e.g. the engine of a moving vehicle) at night. Fig. 5 (b) (c) depict a comparison of thermal imagery and visible imagery in several harsh conditions. In general, it has been demonstrated in Tsimhoni et al. [44] and [45] that at night, the pedestrian detection range of a LWIR camera (165m) is significantly greater than an active NIR camera (59m). Under other adverse conditions, thermal imaging systems are found to be more stable than NIR cameras. The tests conducted in Judd et al. [46] show that LWIR imaging is significantly less affected by fog than NIR cameras. The experiments conducted by Pinchon et al. [47] confirm the advantage of LWIR imagery over NIR imagery in pedestrian detection and demonstrate that the glare caused by oncoming headlamps under fog would not occur in thermal imagery. Nonetheless, the tests in Pinchon et al. [47] show that thermal cameras are unable to detect lane markings or recognize traffic signs, whereas NIR imaging systems can. Thermal cameras are thus more effective but limited in detecting pedestrians or animals in various adverse conditions. Because NIR behaves similarly to the visible spectrum, the NIR imagery provides more contextual information and offers more functions. The active NIR cameras are either adopted as cheaper substitutes for thermal imagery in exterior perception systems, or used for in-cabin driver monitoring systems.

In the automotive industry, thermal cameras have surpassed NIR cameras in market share in night vision systems (NVS). In 2000, General Motors launched the first automotive NVS on the Cadillac DeVille using a LWIR sensor supplied from Raytheon [48]. In 2004, Honda [49] introduced a thermal camera based Intelligent NVS on Honda Legend. From 2005, BMW began to use LWIR cameras in its 7 Series. Peugeot incorporated a thermal camera into its flagship sedan Peugeot 508 from 2018. In 2002, Toyota presented an active NIR NVS in Toyota LandCruiser and Lexus 470, but from 2014 Lexus has decided to discard the NVS in the subsequent generations. When it comes to driver monitoring systems, NIR cameras dominate the market. For instance, in the driver assistant system SuperCruised launched by Cadillac in 2018, an NIR camera is mounted in the instrument panel to monitor whether or not the driver is watching the road in order to raise warnings. In the research field, the situation is similar: most studies on automotive thermal cameras are around pedestrian or animal detection at night. NIR cameras are more investigated for analyzing and monitoring driver’s status. Therefore, the following sections focus the in-cabin applications of NIR cameras and pedestrian/animal detections in LWIR imagery.

1) NIR Cameras in Driver Monitoring Systems: According to the NHTSA (National Highway Traffic Safety Administration), approximately 25% of reported crashes in the U.S. involve a certain form of driver inattention [53]. Distraction and fatigue are the two principal causes of driver inattention. A visual distraction, such as looking away from the front road, is the most common type of distraction. Fatigue can be defined as a subjective feeling of drowsiness caused by physical or mental factors. A driver monitoring system (DMS) utilizes sensors (e.g. image sensor, pressure sensor, etc) to ensure a driver keeping attention on the road, as shown in Fig. 6 (a). A typical DMS usually contains gaze detection and drowsiness detection to warn the driver when an inattention event is detected. Researches (Alhstrom et al. [54], Schwarz et al. [55]) have proved that the DMS could effectively improve safety. In Europe, a general safety regulation has been passed in 2019 to mandate automakers to install advanced safety systems including DMS in new cars on the EU market from 2022. Because an active NIR night vision system is barely perceptible by human eyes and conserves abundant contextual details, it plays a critical role in modern DMS. Face and eye detection and tracking via image processing are usually required as a preliminary step before detecting gaze and drowsiness. Fig. 6 (b) shows an example of detected facial landmarks. In recent years, deep neural networks (DNNs) dominate this domain. For example, Yoon et al. [52] utilize a VGG network for face detection and Park et al. [56] develop a faster-RCNN based eye detection method.

Following the localization of the face and eye regions, additional processing is required to detect drowsiness or distraction. PERCLOS (percentage of eye closure over time) proposed by Dinges et al. [57] is a valid metric for detecting drowsiness that has been used in many studies, e.g. Ji et al. [58], Flores et al. [59], Garcia et al. [60] and Dasgupta et al. [61]. Gaze detection based distraction warning is more complex than drowsiness detection. In the literature, two types of solutions were proposed: 1) geometric approaches and 2) machine learning approaches. The geometric approaches rely

3https://ec.europa.eu/commission/presscorner/detail/en/IP_19_1793
on the 3D gaze estimation via 3D modeling of face/eyes. As in the AttenD algorithm proposed by Ahlstrom et al. [54], the estimated 3D gaze direction is compared with a predefined 3D safe region to detect distraction events. Vicente et al. [50] compute the intersection of the driver’s 3D gaze line and the car windshield plane. An EOR (eyes off the road) event would be triggered when the intersection point lies out of the safe region. Machine learning based methods directly predict a gaze zone from face and eyes image detections, avoiding 3D gaze direction estimation, which can be disrupted by scenario changes. Fig. 6 (c) shows an example of 15 divided gaze zones for gaze classification. Fridman et al. [51] and Naqvi et al. [62] utilize respectively a random forest algorithm on facial landmarks vector, and directly a VGG neural network, to classify the gaze zones, i.e. which zone the driver is looking at. Yoon et al. [52] upgrade this method by using two NIR cameras and residual DNN to improve the accuracy and the robustness. More detailed reviews on gaze detection and DMS could be found in Dong et al. [63] and Akinyelu et al. [64].

Aside from drowsiness and distraction detections, the NIR camera could be used to detect a driver’s vital signs, e.g. pulse rate, respiratory rate, etc, e.g. in Wang et al. [65], Magdalena et al. [66], Wang et al. [67], and Kurihara et al. [68].

2) LWIR cameras for night vision systems: Thermal cameras are particularly suitable for detecting pedestrians and animals at night. Before the era of deep learning, object detection follows a traditional pipeline as: candidate region proposal, feature extraction and machine learning based classification. Popular object detection methods, such as Haar feature-based cascade AdaBoost classifier [69], SVM classifier [70] etc, are still popular in thermal imagery understanding due to less computational cost. Fang et al. [71] manually design features from hotspots in a thermal image to train a SVM classifier to recognize pedestrians. Xu et al. [72] employ a SVM classifier and a mean-shift tracker for pedestrian detection and tracking. Forslund and Bjarkfur [73] present a large animal thermal image dataset gathered over an 8-year period of driving in various locations. A cascade AdaBoost classifier is applied for animal detection and warning driver assistance systems. Savasturk et al. from Daimler [74] find significant benefits in vehicle detection by combining RGB stereo images with monocular thermal images. In a recent work [75], Chen et al. discover that by feeding CCF (Convolutional Channel Features) to a cascade AdaBoost classifier, LWIR camera can outperform stereo vision in pedestrian detection.

Entering into the era of deep learning when the CNNs (Convolutional Neural Networks) sweep all the computer vision benchmarks, there is no exception in thermal image processing. Kristo et al. [76] benchmark several popular object detectors, including Faster R-CNN [77], SSD [78] and YOLOv3 [79], that are retrained on a thermal image dataset for a surveillance system. YOLOv3 has been found to be significantly faster than other methods while still achieving comparable performance to the best. Dai et al. [34] propose a TIRNet for pedestrian detection by modifying the SSD detector. The performance of TIRNet is reported better than YOLOv3 based on their annotated dataset and the KAIST dataset. Fig. 7 (a) shows several detection results of TIRNet. A large-scale thermal pedestrian dataset SCUT is presented in Xu et al. [80]. Based on this dataset, the authors provide a detailed comparison between widely used detectors. Tumas et al. [81] present a ZUT dataset containing vehicle odometry and weather measures. Other than re-implementing the CNN detectors for thermal image processing, Grimming et al. [82] dive deeply into a thermal camera’s physical characteristics, and studied the relation between MTF (modulation transfer function), NETD, and the performance of fast R-CNN object detector [83].

In practice, fusing thermal and RGB images to deal with complex conditions is still indispensable for the majority of engineering systems [84]. The fusion of DNNs from multispectral data could be done in different stages, as illustrated in Fig. 7 (b). The strategies can be roughly divided by the time to fuse: input fusion, early fusion, halfway fusion, late fusion and score fusion. Late fusion offers the flexibility to directly fuse existing detectors inferring in parallel. Choi et al. [85] and Park et al. [86] fuse this way two CNNs for proposal generation on color and thermal streams. With more modalities, Humbolt-Renaux et al. [87] investigate the late fusion for multispectral people detection from YOLO detectors, as well as Takumi et al. [88] from RGB, NIR, MIR and LWIR images. Other authors, e.g. Wagner et al. [89], Liu et al. [90], Li et al. [91], compare fusion schemes for pedestrian detection. The findings show that halfway fusion is superior to other approaches. As a result, the halfway fusion has become the default fusion strategy in CNN based multispectral image understanding, as demonstrated by Li et al. [92], Guan et al. [93], Zhang et al. [94] and Yadav et al. [95].
Another trend is looking for new neural network modules. Konig et al. [96] use the faster-RCNN framework but with a new region proposal network (RPN) based on LWIR and RGB semantic segmentation. Zhang et al. [94] propose a “Cyclic Fuse-and-Refine” module to optimize the complementary and consistency of multispectral features. In Li et al. [91] and Guan et al. [93], illumination detection modules are proposed to dynamically assign the weights of multispectral features under a halfway fusion architecture. Zhang et al. [97] propose an Aligned Region CNN (AR-CNN): a neural network module that compensates spatial misalignments of the features extracted from thermal-RGB image pairs. Dasgupta et al. [98] extend the halfway fusion architecture with a multimodal feature embedding module (MuFEm) and a CRF-based Spatial-Contextual feature aggregation module.

Apart from object detection, thermal cameras could segment road segments as well, e.g. in Pelaez et al. [99], Yoon et al. [100], Humblot-Renaux et al. [87].

IV. RANGE-GATED CAMERA

To enhance the imaging quality under harsh conditions, range-gated imaging technique was firstly proposed in 1960s [101] and has been applied in night vision system [102], submarine vision [103]. In recent years, range-gated cameras have gained popularity because of their resistance to adverse conditions [104].

A. Principles

A range-gated camera is an active imaging system in which an illuminator transmits pulsed light, and an image sensor is precisely synchronized to image the reflected lights within certain defined “gates”. A general principle of a range-gated imaging system is shown in Fig. 8. In the illuminator module, light pulses are emitted to illuminate the environment within the lens’s field-of-view. Parts of the transmitted lights will be reflected by the surfaces of the objects and then be captured partially by the receiving optics. Because the objects are at different ranges, the reflected photons are captured at different times. Unlike conventional cameras’ exposure methods (global shutter or rolling shutter), a gated camera employs several gate functions to expose the photons arriving at different times. Therefore, only the light arriving within the right timing window contributes to the final image. Usually, the exposure gates are very short: in the order of 0.01 - 2µs. As the example in Fig. 8, three programmed gated functions generate three image slices containing objects at different ranges. The final image is obtained by merging those image slices. The components are introduced as follows:

Illuminator is triggered by the gating signals from a controller. Owing to narrow spectral width and high peak power, the laser is preferred over other kinds of lights. Different laser wavelengths ranging from visible, NIR to SWIR wavelengths could be applied. The NIR laser is popular because of its maturity and cost. For instance, 808nm laser is used in David et al. [105] and Spooren et al. [106]. When considering better penetration in long-distance through fog or smoke, SWIR laser is preferred because it can achieve much higher transmission power while still meeting the eye safety standards. In [102], a range-gated imaging system based on a Nd YAG laser at 1571nm reaches a 10km detection range. Similarly, in Baker et al. [107], a range-gated SWIR (1527nm) camera successfully penetrates heavy rains and detects obstacles 10km away.

Gated image sensor: The gated image sensors can perform multi-integration to generate a merged image by using gated signals. Due to the extremely short integration time, the gated image sensor has to be highly efficient. In Spooren et al. [106], a gated RGB-NIR image sensor with high NIR quantum efficiency (40%) is built. In Rutz et al. [108], a high-gain avalanche photodetector (APD) array containing 640 × 512 InGaAs pixels is coupled with a SWIR laser transmitter. When operated in Geiger mode, the APDs become single-photon avalanche diodes (SPADs), meaning that even a single photon could trigger the avalanche effect. Burri et al. [109] present a 512 × 128 pixel CMOS SPAD sensor capable of operating within an exposure window as small as 4ns. In Morimoto et al. [110], a 1M pixel CMOS SPAD image sensor is built for 3.8ns gating time.

B. Applications of in ADAS/AD

The range-gated cameras, as active sensors, are better suited to low-light conditions, such as a country road at night. Furthermore, owing to its imaging mechanism, only photons received at appropriate times are utilized for imaging. Such an attribute has two advantages: (1) No blooming effect when the photons from highly reflective objects do not
fall within the sampling range. For example, the oncoming vehicles’ headlamps have almost no impact on range-gated images. (2) Resistance to backscattering environments, such as fog/rain/smoke. A key parameter deciding the quality of a range-gated image is the modulation contrast:

\[
Contrast \simeq \frac{I_{\text{target}} - I_{\text{background}}}{I_{\text{target}} + I_{\text{background}} + 2I_{\text{bsc}}} \tag{4}
\]

Where \(I_{\text{target}}, I_{\text{background}}, I_{\text{bsc}}\) are the luminance of the target, background, and due to backscattering effect. Hence, the image quality in backscattering condition is defined by the strength of \(I_{\text{bsc}}\), which can be calculated as:

\[
I_{\text{bsc}} = \int_{2\gamma R_{\text{on}}}^{2\gamma R_{\text{off}}} PGe^{-\gamma x^2} \frac{F_n\theta^2 X^2}{2F_n^2\theta^2 X^2} dX \tag{5}
\]

where \(R_{\text{on}}, R_{\text{off}}\) define a range interval during one exposure. \(G\) is the backscatter gain, \(\gamma\) is the atmospheric attenuation coefficient, \(\theta\) is the laser beam divergence, \(P\) is the laser power, \(F_n\) is the speed of the lens, and \(X\) is the integration variable. Comparing with a conventional camera that all the backscattering photons are counted, a range-gated camera only perform the photon integration during a very short opening time, i.e. between \(R_{\text{on}}\) and \(R_{\text{off}}\), so that a higher contrast defined in Eq. 4 can be achieved. A comparison between a RGB camera and a range-gated camera in a fog environment is shown in Fig. 9. Walz et al. [112] benchmark multi-model sensors in a well controlled artificial fog chamber. Both the quantitative and qualitative results show the superiority of range-gated cameras in such harsh conditions. Owing to the excellent performances in adverse conditions, the range-gated camera has the potential to be a strong competitor to infrared cameras, and has gained recognitions in recent years. On the industry side, [113] first apply a NIR range-gated camera to aid driving at night. Grauer et al. [114] and [115] present a high resolution (1.2M pixel) range-gated camera based on NIR VCSEL laser (808nm) and a gated CMOS image sensor. This sensor is suitable for use in active safety systems such as vulnerable object detection, forward collision warning, lane departure warning, traffic sign detection, etc.

From 2017, a series of works around range-gated camera images are developed within the EU-founded DENSE project 4. Supported by this project, the DENSE dataset 5 containing multi-model sensors (a range-gated camera, a RGB stereo

---

4 https://www.dense247.eu/
5 https://www.uni-ulm.de/en/in/driveu/projects/dense-datasets/
A LWIR camera, and a LiDAR) is released to the public. The dataset covers snow, rain, urban and sub-urban scenarios. The DENSE dataset is further annotated in Julca-Aguilar et al. [116] as Gated3D dataset, in which more than 100K objects in 4 classes are manually annotated over 12997 image frames. Based on these datasets, Tobias et al. [117] present a deep neural networks (DNN) named as "gated2depth", which can estimate the depth of each pixel in the range-gated camera. The proposed CNN architecture utilizes all the three slice images. Walz et al. [112] extend gated2depth by incorporating aleatoric uncertainties into the pixel-wise depth estimation. Bijelic et al. [104] propose a CNN for adaptively fusing LiDAR, RGB camera, gated camera and radar features in an entropy estimation framework (higher entropy indicates more confidence). A delicate feature exchange network is designed to dynamically allocate the best features for each sensor. To explore the implied range information in the slice images, Julca-Aguilar et al. [116] propose a CNN for 3D object detection. The proposed CNN is tailored to the temporal illumination cues from the three image slices. Based on the Gated3D dataset, they demonstrated that using temporal cues from a range-gated camera, the 3D object detection results outperform a pure RGB based detection method.

V. POLARIZATION CAMERA

A. Principle

According to Sec. II-A, light passes through a medium as a transverse wave, i.e. oscillating perpendicularly to the direction of propagation, that consists of an oscillating electric field and a magnetic field. For computer vision applications, only the electric field is considered. Polarization is a fundamental and distinct property that describes the orientation of the light oscillation [120]. There are in general three kinds of polarized light: totally polarized (linear, circular or elliptic), partially polarized and unpolarized. The majority of the light sources, e.g. the sun, streetlamps, emit unpolarized light, i.e. it vibrates randomly in all directions.

Although most natural light is unpolarized, it can be converted to polarized light through the reflection from certain surfaces. In an ideal situation when the incident angle of unpolarized light is the angle of Brewster, according to Fresnel equations [121], the reflected light is linear polarized (as shown in Fig. 10 (a)). Otherwise, it would be partially polarized. Reflections from most flat surfaces are partially polarized as a function of incident angle. A more controllable way to obtain polarized light is to use a polarizer, which is an optical filter that passes only specific polarized light while blocking light from other polarizations, as shown in Fig. 10 (b).

A concise representation of polarized light is the Stokes vector $\mathbf{S}$ [122], consisting of 4 parameters: $\mathbf{S} = [S_0, S_1, S_2, S_3]$. $S_0$ is the total light intensity, $S_1$ and $S_2$ roughly represent the degree of linearly polarized light ($S_1$ stands for horizontal or vertical linear polarization, $S_2$ stands for 45° or 135° linear polarization). $S_3$ stands for ellipticity, which is usually ignored in applications.

$$S_0 = I_0 + J_90 = I_{45} + I_{135}$$

$$S_1 = I_0 - J_90, \quad S_2 = I_{45} - I_{135}$$

(6)

where $I_0$, $I_{45}$, $J_90$ and $I_{135}$ are the optical intensities at the corresponding polarization direction, i.e. $0°$, $45°$, $90°$ and $135°$. Other important physical properties, e.g. angle of polarization (AoP) and the degree of polarization (DoP) can be inferred from the Stokes vector as:

$$AoP = \frac{1}{2} \arctan\left(\frac{S_2}{S_1}\right), \quad DoP = \frac{\sqrt{S_1^2 + S_2^2}}{S_0}$$

(7)

Varying between 0° and 180°, AoP represents the predominant axis of the light vibration. DoP is the ratio of the intensity of the polarized portion to the total intensity. For instance, a linearly polarized light has a DoP of 1, natural light usually has DoP between 0 to 0.5.

Creating a practical and convenient polarimetric imaging system is not a easy work. In early research, Morel et al. [123] made a polarization camera by manually rotating a polarizer in front of a normal camera. Three images are taken at different rotating angles of the polarizer to determine the Stokes vector for each pixel. In Wolff et al. [124], a polarizing beam splitter is placed in front of 2 cameras so that the reflected and the transmitted beams are utilized to compute the polarization of each pixel. However, those methods either require a special environment for imaging, or are too expensive.

Powered by on-chip polarizer technology, modern image sensors can simultaneously acquire polarization and color information through a single shot. For instance, inside SONY’s Pregius IMX250 CMOS sensor (as shown in Fig. 10 (c)), a Polarization Filter Array (PFA) composed of four various angled micro-polarizers ($0°$, $45°$, $90°$, $135°$) is placed on top of the CFA and photodiodes. The Stokes vector (as in Eq. 6) and RGB vector for each pixel can be interpolated by using a special demosaicing process afterward. Such snapshot technology has a price advantage that it has been employed in many computer vision studies.

B. Applications in ADAS/AD

Polarization cameras are not yet commercialized for automotive usages. Nevertheless, as the snapshot P-RGB image sensors (e.g. SONY IMX250, IMX253) become more popular, more researchers are beginning to investigate the potential benefits of light polarization. Current research focuses on image enhancement, object detection and semantic segmentation.

As discussed in Sec. II-C, the specular reflection, high contrast regions, and adverse weather would degrade the image quality. Although many intensity-based solutions (e.g. Li et al. [125], and Wang et al. [126]) could alleviate these issues, polarization cameras offer a new perspective. Wang et al. [127] utilize a polarization camera to remove specular reflection because the DoP of the specular reflection part is much larger than the part of diffuse reflection, when an unpolarized light beam is reflected. Polarization cameras can also achieve high dynamic range (HDR) imaging to solve the over/under-saturation in high contrast conditions. As proposed by Wu et
al. [128], the 4 micro-polarizer patterns have similar effects as 4 different exposure times. Therefore, by using multiple polarization images at known pixel-specific exposure times, the irradiance maps can be estimated and hence construct an HDR image.

Polarimetric images provide physical properties of the object, such as surface material and roughness, which can be utilized as a complement to traditional RGB image based object detection and segmentation. Wang et al. [131] implement a feature selection process in polarimetric images and discover that the AoP is the most informative polarization feature. Then, for car detection, the AoP features are incorporated with a deformable part based models (DPM). The experimental results demonstrate that polarization features significantly reduce the false detection rate. Adding the polarization features to an object detection DNN, Blin et al. [132] and [133] show that car detection results under adverse weather conditions could be improved by 20% - 50%. In addition, a new dataset PolarLITIS [134] containing RGB and polarimetric images under fog conditions was released to evaluate the performance gain of object detection from polarization information. The experiments in Blanchon et al. [135] and Xiang et al. [136] both find that the semantic segmentation for car and windows is largely improved thanks to the polarization features.

VI. EVENT CAMERA

In dynamic and unpredictable environments, traditional cameras would give blurry images or under/over-exposed images. The neuromorphic vision sensor is a good choice for a robust perception system. A general survey on event cameras is given in [137], and a tutorial aiming at some common processing methods applied for autonomous driving is given in [138]. This section is complementary to these papers in providing a review of event vision for driving applications.

A. Principles

The event camera is also called address-event representation silicon retina, neuromorphic, or retinomorphic camera, because it is inspired by eye retina, as described by [129]. In the retina, the fundus of the eye, are located the cones and rods cells, that are sensitive to light, followed by layers of neurons. Photosensitive cells convert light into electric signal transmitted to nerve cells. Some signal exchanges occur from each photosensitive cell up to two bipolar ganglion cells: when activated, the first one represents ON pulse whereas the second one represents OFF pulse. In summary, ON cell activates when a spatiotemporal brighter change in contrast occurs, OFF cell activates when a spatiotemporal darker contrast change occurs. The brain is able to interpret these voltage spikes to give to us our sight sense. This process leads to the following advantages: Independence from absolute light level: It can be seen as an automatic gain control from the retina and allows vision capabilities for a very wide range of brightness. Lightweight data encoding for fast transmission. Spikes are emitted continuously to the brain, avoiding the need to encode absolute intensities, and giving a high temporal resolution.

An event camera is designed to imitate the retina by bionomic pixel circuits (as shown in Fig. 11 (a)), and hence inherit these advantages. Pixel outputs of an event camera are independent, they represent signed spikes as long as the photosensor observes a log-intensity difference above a threshold. The rate of following spikes of the same sign is an indication of the brightness change speed. Then, a stream of events is a sequence of timestamped signals, where each signal represents a positive or negative pulse (that is respectively, a state change to be more or less bright) for one or several points of the matrix sensor. An event camera does not stream full image frames in the way a conventional camera does at a given framerate. It acts in an asynchronous way with a very high temporal resolution and low latency, in an order of microseconds. The output difference between both sensors is shown in Fig. 11 (b). Similar to conventional image sensors, event sensors are made of Silicon and are sensitive to visible and NIR light. On the contrary, event cameras are often made without IR cut filter in order to gather more light. However, the use of specific wavelength filters may be necessary for certain applications. Modern neuromorphic cameras reach HD resolution, such as: Prophesee Gen4 CD (1280×720 pixels) [139], Samsung DVS-Gen4 (1280×960 pixels) [140], Celepixel gallery (1280×800 pixels) [141].
Some event cameras (e.g. iniVation DAVIS346 \(^6\)). CeleX-V incorporates additional circuits in order to simultaneously output conventional images (monochrome in most cases) and sensed events. Such design gives the advantage of data fusion at exact superimposition, while at the expense of increasing noises caused by residual currents brought by those additional circuits. Rare event cameras are able to output both RGB events and frames, as iniVation DAVIS346B-Color\(^7\), which includes a Bayer filter array to estimate RGB channels. Sample data of RGB event camera is available through the Color Event Camera Dataset (CED) \([142]\).

**B. Advantages**

Event cameras are bio-inspired passive sensors that try to imitate millions years evolution of sight sense. General advantages of event camera are stated by \([137]\): *Microsecond temporal resolution* for detection and timestamp. A direct consequence is the ability to always avoid motion blur as it exists for conventional cameras. Furthermore, the event camera operates at sub-millisecond latency, which is approximately equivalent to a virtual \(> 1000 FPS\) frame-based camera. *Low power consumption*, in the order of \(10mW\) to \(100mW\) for typical event cameras, while usually between \(1W\) and \(3.5W\) for industrial RGB cameras. *Broad dynamic range*: an event camera’s dynamic range can easily reach \(>120dB\) without a special design. In contrast, a normal RGB camera needs a dedicated pixel design to boost its dynamic range from typical \(60 - 70dB\) to \(110dB\).

All these advantages are desirable for intelligent vehicles: very high temporal resolution allows for detection of fast-moving elements of the scene; very low-latency is important for safety-critical applications; very high dynamic range allows to perceive in challenging lighting conditions. Event camera capabilities in driving scenes are illustrated in Figure 12.

\(^6\)http://inivation.com/wp-content/uploads/2020/09/DAVIS346.pdf

\(^7\)http://inivation.github.io/inivation-docs/Hardwareuserguides/User_guide_-_DAVIS_USB3_development_kit.html

**C. Data representation and processing**

Unlike frame cameras, neuromorphic cameras output stream of events \(E = \{e(x, y, t, p)\}\) that each event \(e(x, y, t, p)\) encodes position \((x, y) \in \mathbb{N}^2\), timestamp \(t\), polarity of the brightness change \(p \in \{-1, +1\}\). In signal processing point of view, an event \(e(x, y, t, p)\) can be considered as a continuous function using diracs \(e = p \cdot \delta(\xi - x, \upsilon - y) \cdot \delta(t - \tau)\) where \((\xi, \upsilon) \in \mathbb{R}^2\) represents 2D spatial positioning in pixel array and \(\tau \in \mathbb{R}^+\) represents continuous running time. There are two ways of processing event flows: asynchronous processing when an event arise, that is event-by-event processing, or accumulation of events within a temporal window, i.e. process them as an array or as a tensor.

**Event-by-event processing**: it is a natural way to keep the raw asynchronous and sparse event(spike) flow, whereas current computers are not designed for spikes processing. Standard processors architectures (CPUs and GPUs) are good to process dense arrays of data, but are not able to process irregular flows of independent events at a very high rate. Specific biologically inspired hardware is designed to efficiently process event-by-event, such as ROLLS processor \([143]\), IBM TrueNorth chip \([144]\), and Intel Loihi chip \([145]\). These spike processors are particularly interesting since they open the door for hardware SNN (spiking neural networks) with low power consumption. SNNs are designed to imitate brain neurons and are the most popular and direct way to process event-by-event flows. With or without specific hardware, some early investigations about SNNs have been done. However, as a single event gives insufficient information for understanding, new events are used iteratively to update a system’s state. While some methods apply standard optimization techniques or filters (such as \([146]–[148]\)), most of them integrate asynchronous events in artificial neural networks. SNNs have already been proposed for many applications. For example, \([149]\), \([150]\) for stereo depth estimation, \([151]\) for classification, \([152]\) for optical flow, \([153]\) for background motion separation, \([154]\), \([155]\) for heading estimation and loop closure detection, \([156]\), \([157]\) for robotic control, \([158]\) for target following, \([159]\) for...
collision avoidance (for drone).

Events binning: It is generally more practical to process batches of incoming events rather than processing individual events. Usually, successive events are gathered and compressed into a dense array or a tensor that is similar to an image frame. Both arrays and tensors can be efficiently processed by standard computer hardware. Binning events cause additional latency. However, this drawback is moderate and acceptable as the general advantages are still kept. There are two strategies of events binning: via a time window, or via a queue of a fixed number of events. The usage of time windows is easy and common, while can lead to accumulated arrays without or too many events. Hence, the sampling time should be tuned accordingly and gives a synchronous process. Zou et al. [160] use adaptive accumulation time (making the method asynchronous), Rebecq et al. [161] use overlapping pairs of time windows, and Joubert et al. [162] combine time windows of different lengths. The study [163] contains a performance analysis of various time window durations for a classification task. The strategy of binning by a fixed number of events is used in [164]–[167]. It allows for keeping accumulated representations with similar appearances (same density of events) and for asynchronous process. Nevertheless, the following operations should be fast enough when huge quantities of events arrive in a short time.

After accumulating enough events, the next issue is encoding, i.e. extracting effective event attributes. Several handcrafted encoding methods have been proposed in the literature. For example, leveraging “frequency encoding” representation from Chen et al. [168], where a standard YOLOv3 CNN architecture [79] is utilized for pedestrian detection. Chen et al. [169] also get the best results with the “frequency encoding” among other encoding schemes for driver monitoring applications. Perot et al. [170] test different accumulation and encoding strategies for object detection, with the best results using the “discretized event volume” representation from Zhu et al. [171]. Besides human-designed features, generalized expression can also be learned automatically in an end-to-end manner. Tulyakov et al. [172] model events as a stream of sparse 3d data points, and then apply a MLP (Multi-Layer Perceptrons) to learn an optimal encoding for a stereo-matching problem. Experimental results show that the learning-based encoding is better than the best handcrafted approach. Cannici et al. [173] propose specific LSTM (Long Short-Term Memory) recurrent modules as a flexible way to learn task-dependent event-surfaces, and show better performance in optical flow estimation. Li et al. [174] apply a SNN to encode events and generate visual attention maps for further fusion with frame images, in an object detection framework.

At last, event accumulation can be motion compensated with a fast algorithm, typically using a joint IMU inside event camera [161]. This guarantees for event accumulation with no blur effect in case of a long accumulation time.

D. Applications in Autonomous Driving or ADAS

The event camera is especially useful for systems running with real-time interactions, non-controlled enlightenment conditions, and low latency. In this paper, we focus on their application in the autonomous driving field.

1) Dataset in driving scenes: To apply the event cameras in autonomous vehicles, large and well-annotated datasets are indispensable. The neuromorphic vision community is very active in it. Because event cameras are still in early stages, many published datasets (e.g. MVSEC [175], DDD17 [176], etc.) in recent years are still in low image resolution (e.g. less than 640×480) due to hardware limits. The first HD event camera released in public is the CelePixel CeleX-V in 2019 [141]. Larger resolution benefits further object detection range and better recognition for small objects, while poses challenges for computation capability because of huge event flows. We expect to see more and more HD event camera datasets as Perot et al. [170] appear in public. The published datasets are for various purposes, such as target detection...
2) Object detection: Object detection is a traditional but critical topic for autonomous driving systems. Since the event camera is a new sensor, labeled event datasets are scarce. Leveraged by the pseudo-labels warped from frame images in DDD17 dataset, Chen et al. [168] apply several popular CNN object detectors in event camera images, and achieved good results in motion-blurring scenarios. The PKU-DDD17-CAR dataset, i.e. the annotation of DDD17 by Li et al. [174], is used by Cao et al. [178] to detect vehicles. Hu et al. [179], [180] augment respectively parts of a day and a night sequence of MVSEC dataset [175] with car annotations for DNN training. Except for events-only object detection, another direction is to fuse frame images for better performance. Li et al. [174] apply a SNN for the event stream to generate attention maps that feed to a CNN concatenated to standard frames, as in an early fusion scheme. Cao et al. [178] fuse events and frames at different encoding levels from parallel heads using feature attention gate components. Hu et al. [179], [180] illustrate proposed grafted networks and events synthesis from video frames with a car detection use case. Pedestrian detection is also important and is explored in [181]–[184], in which individual datasets are utilized according to specific cameras. Chen et al. [181] compare different accumulation methods coupled with early fusion and late fusion schemes. In Jiang et al. [182], events and frames data channels are both fed into different CNNs and then fuse multiple confidence maps to achieve good pedestrian detection. Chadera et al. [183] implement a BNN (binary neural networks) on FPGA for fast detection. Wan et al. [184] propose a Pedestrian-SARI dataset and alternative events representations for asynchronous CNN detection. Lane extraction problem is investigated in Cheng et al. [177], in which a DET dataset, labeled lane markings in HD event camera, is released to be public. Meanwhile, several popular CNN-based lane extraction algorithms are benchmarked and the results show good performances. A more general object detection method and an annotated dataset are proposed in Perot et al. [170], where a CNN combined with a LSTM is used to keep detections when movements stop. The proposed method is evaluated for HD events road scenes, released in their 1 Megapixel Automotive Detection Dataset.

3) Motion segmentation: motion segmentation or moving object detection by an event camera is more convenient than a conventional camera. This topic is addressed in [185]–[188]. In general, those approaches compensate first camera motion as background movement, as it is likely to cause the most prominent number of events. Then, moving objects are segmented through different clustering strategies. For instance, Mitrokhin et al. [185] group events into clusters via morphological operators, then tracks the multiple moving objects. Stoffregen et al. [186] warp the events several times to cluster moving objects. Zhou et al. [187] cluster the objects via graph cut on linked space-time event graph. [188] cluster the objects with split and merge strategy and track grouped events. Monda et al. [189] don’t consider background motion. Instead, moving objects are segmented from a fixed event camera flow, and then grouped by a k-NN graph method.

4) Driver monitoring system: Currently, few investigations have been done with event cameras to monitor driver status. Chen et al. [190] focus on drowsiness detection, and compare some classification algorithms on their event dataset. Provided with a new dataset, Chen et al. [169] compare different CNN architectures and events accumulation schemes for driver drowsiness detection, gaze-zone recognition, and hand-gesture recognition.

E. Remaining Challenges

Although the attributes of event cameras are attractive, they are still quite young, and hence suffer several restrictions for wide applications. The first restriction involves the optimal performance in complex enlightenment scenarios, for example, the camera biases and noise. Biases need to be carefully tuned to achieve optimal perception according to the conditions (scene brightness and dynamics, ambient temperature, admissible noise, etc). Tuning the event camera’s parameters is not a straightforward task, as there are too many correlated parameters to adjust. Usually, some general tries are required at first to correct the parameters. Details on how to control an event camera are given in Delbruck et al. [207]. Other constraints concern the sensing characteristic. The most typical issue is the relative static object, as illustrated in Fig. 12 (c), because the front car and ego-vehicle are both stopped, the event camera barely perceives the front car. Fortunately, such a problem could be overcome by applying RNN in object detection Perot et al. [170]. Irregular data bandwidth is another constraint caused by a huge amount of events generated at instants. Unlike a fixed bandwidth for a frame camera, the bandwidth of an event camera could reach the limits of a vehicle’s onboard network capabilities, causing network jamming or package losses. To handle this problem, Khan et al. [208] propose an efficient compression algorithm. Another issue is the shadows (in Fig. 12 (d)), the shadow of a traffic sign on the ground may generate a false alarm.

VII. Conclusion and Future Works

Although RGB cameras have cost advantages and are widely applied in vehicles, several inherent limitations impede the development of a better autonomous driving system working in larger ODDs. To overcome these drawbacks, other kinds of sensing modalities are emerging. In this paper, we’ve reviewed several emerging vision sensors as complementaries for conventional RGB cameras: infrared, range-gated, polarization, and event cameras. Some of them have been already integrated into production cars, such as NIR and LWIR cameras. While some of them are still in early exploration stages, such as the polarization, event and range-gated cameras. With the additional sensing modalities, an autonomous vehicle is expected to be operated in larger ODDs (e.g. at night or through rainy days). Most of the perception algorithms for the introduced sensors are similar to processing RGB images that RGB channels are replaced by infrared or polarization channels. For range-gated and event cameras, since their imaging principles are different, specifically designed algorithms are
powerful algorithms to utilize their imaging advantages. A fusion between those sensors and RGB cameras is more practical in designed to leverage their unique imaging property. A fusion between those sensors and RGB cameras is more practical... summarized in Fig. 7 (b). In the future, we believe that the reviewed sensing technologies will have better maturity, lower price, and more powerful algorithms to utilize their imaging advantages.

REFERENCES

[1] C. Urmson et al., “Autonomous driving in urban environments: Boss and the Urban Challenge,” Journal of Field Robotics, vol. 25, pp. 425–466, 2008.

[2] “Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles,” SAE International, Tech. Rep., 2018.

[3] Y. Li and J. Ibanez-Guzman, “Lidar for autonomous driving: The principles, challenges, and trends for automotive lidar and perception systems,” IEEE Signal Processing Magazine, vol. 37, pp. 50–61, 2020.

[4] S. M. Patole, M. Torlak, D. Wang, and M. Ali, “Automotive radars: A review of signal processing techniques,” IEEE Signal Processing Magazine, vol. 34, pp. 22–35, 2017.

[5] J. Van Brummelen, M. O’Brien, D. Gruyer, and H. Najjaran, “Autonomous vehicle perception: The technology of today and tomorrow,” Transportation Research Part C: Emerging Technologies, vol. 89, pp. 384–406, 2018.

[6] E. Yurtsever, J. Lambert, A. Carballo, and K. Takeda, “A survey of autonomous driving: Common practices and emerging technologies,” IEEE Access, vol. 8, pp. 58443–58469, 2020.

[7] C. J. Cleveland and C. Morris, Handbook of Energy. Elsevier, 2013.

[8] “Standard tables for reference solar spectral irradiances: Direct normal and hemispherical on 37° tilted surface,” American Society for Testing Materials, Standard, 2012.

[9] I. D. Boullough et al., “An investigation of headlamp glare: Intensity, spectrum and size,” 2003.

[10] UNECE, “Regulation of no 112 of the economic commission for europe of the united nations (unece),” Tech. Rep., 2014.

[11] P. Duthon, M. Colomb, and F. Bernardin, “Light transmission in fog: The influence of wavelength on the extinction coefficient,” Applied Sciences, vol. 9, pp. 2843–2854, 2019.

[12] K. Murari, R. Etienne-Cummings, N. Thakor, and G. Cauwenberghs, “Which photodiode to use: A comparison of cmos-compatible structures,” IEEE Sensors Journal, vol. 9, pp. 752–760, 2009.

[13] A. Carrasco-Casado and R. Mata-Calvo, Space Optical Links for Communication Networks. Springer International Publishing, 2020, pp. 1057–1103.

[14] K. Weidt, D. Schroeder, and W. Stechle, “Optimization of automotive color filter arrays for traffic light color separation,” in Color and Imaging Conference, 2020.

[15] M. Roser and P. Lenz, “Camera-based bidirectional reflectance measurement for road surface reflectivity classification,” in IEEE Intelligent Vehicles Symposium, 2010.

[16] P. Duthon, F. Bernardin, F. Chausse, and M. Colomb, “Methodology used to evaluate computer vision algorithms in adverse weather conditions,” in Transportation Research Procedia, 2016.

[17] R. Thakur, Infrared Sensors for Autonomous Vehicles, in Recent Development in Optoelectronic Devices. IntechOpen, 2017.

[18] OmniVision. (2019) Rgb-ir technology. [Online]. Available: https://www.ovt.com/purecel-pixel-tech/rgb-ir-technology/faqs

[19] M. Brown and S. Susstrunk, “Multispectral sift for scene category recognition,” in IEEE/CVF International Conference on Computer Vision and Pattern Recognition (CVPR), 2011.

[20] C. Fredembach and S. Susstrunk, “Colouring the near-infrared,” in Proceeding of 16th Color and Imaging Conference, 2008.

[21] J. R. Dean, “Using near-infrared photography to better study snow microstructure and its variability over time and space,” Master’s thesis, Boise State University, 2016.

[22] OmniVision. (2019) Nyxel® technology generation2. [Online]. Available: https://www.ovt.com/purecel-pixel-tech/nyxel-technology-generation-2

[23] E. de Borniol et al., “High-performance 640 x 512 pixel hybrid ingaas image sensor for night vision,” in Proc. SPIE 8535, Infrared Technology and Applications XXXIII, 2012.

[24] Z. Chen, X. Wang, and R. Liang, “Rgb-nir multispectral camera,” Optics Express, vol. 22, 2014.

[25] Y. M. Lu, C. Fredembach, M. Vetterli, and S. Susstrunk, “Designing color filter arrays for the joint capture of visible and near-infrared images,” in 16th IEEE International Conference on Image Processing (ICIP), 2016.

[26] C. Park and M. G. Kang, “Color restoration of rgb multispectral filter array sensor images based on spectral decomposition,” Sensors, vol. 16, no. 7, 2016.

[27] O. Skorka, P. Kane, and R. Ipsasou, “Color correction for rgb sensors with dual-band filters for inabin imaging applications,” in Electronic Imaging, Autonomous Vehicles and Machines Conference, 2019.

[28] B. Geelen, N. Spooren, K. Tack, A. Lambrechts, and M. Jayapala, “System-level analysis and design for rgb-nir cmos camera,” in Proc. SPIE 10110, Photonic Instrumentation Engineering IV, 2017.

[29] M. Dummer, K. Johnson, S. Rothwell, K. Tatata, and M. Hibbs-Brenner, “The role of vcsels in 3d sensing and lidar,” in Proc. SPIE 11692, Optical Interconnects XXI, 2021.

[30] R. H. Vollmerhausen, R. G. Driggers, and V. A. Hodgkin, “Night illumination in the near- and short-wave infrared spectral bands and the potential for silicon and indium-gallium-arsenide imagers to perform night targeting,” Optical Engineering, vol. 52, 2013.

[31] F. Rutz et al., “Ingaas infrared detector development for swir imaging applications,” in Proceedings Volume 8896, Electro-Optical and Infrared Systems: Technology and Applications X, 2013.
## TABLE II
Event camera datasets with driving scenes. Top part lists low resolution datasets ($< 1280 \times 720$px), bottom part lists high resolution datasets ($\geq 1280 \times 720$px).

| Name | Pixel resolution | Other modalities | Aimed problems | Size | Annotations | Location | Year |
|------|------------------|------------------|----------------|------|-------------|----------|------|
| PRED18 [165] (includes previous PRED16 [164]) | $240 \times 180$ | Grey frames* | Mobile target following | 1.25h | prey size, prey position | Northern Ireland | 2018, 2016 |
| DDD20 [195] (includes previous DDD17 [176]) | $346 \times 260$ | Grey frames*, IMU*, Car data, GNSS | Vehicle control | 39h + 12h | - | USA, Switzerland, Germany | 2020, 2017 |
| PKU-DDD17-CAR* [174] Ev-Seg† [196] N-Cars [197] | $304 \times 240$ | - | Classification | 24K samples | "Car", "Background" | unknown | 2018 |
| MVSEC [175] | $346 \times 260$ | 2 cameras, IMU, LiDAR, GPS, Motion capture | Depth, Localisation | 1h | Depth | USA | 2018 |
| MVSEC-OF*† [180] | | | Optical flow | Partial seq. "outdoor_day2" | "Car" | | 2020 |
| MVSEC-NIGHTL21† [180] | | | Detection | Partial seq. "outdoor_night1" | "Car" | | 2021 |
| Slasher dataset [199] | $346 \times 260$ | Grey frames* | Steering, Radio localisation | Vehicle control | 2 sequences | - | Swiss | 2019 |
| Event Camera Driving Sequences [200] | $640 \times 480$ | RGB camera | Frames reconstruction | 40 sequences | - | - | Swiss | 2019 |
| CED [142] | $346 \times 260$ | RGB event cam, Grey frames* | Color frames reconstruction | 50min | - | unknown | 2019 |
| Pedestrian Detection Dataset [201] | $346 \times 260$ | RGB event cam | - | Detection | 12 recordings | "Pedestrian" | China | 2019 |
| EDDDD [190] | $346 \times 260$ | - | Driver monitoring | 260 sequences | Drowsiness | China | 2020 |
| NeuroIV§ [169] | $346 \times 260$ | RGB event cam, Depth maps, NIR frames | Driver monitoring | 27K samples | Drowsiness, Gaze-zones, Hand-gestures | China | 2020 |
| GAD Dataset [202] | $304 \times 240$ | - | Detection | 39h | "Car", "Pedestrian" | France | 2020 |
| Brisbane Event VPR [203] | $346 \times 260$ | RGB event cam, RGB frames*, IMU*, GPS | Visual place recognition | 8km | Landmarks | Australia | 2020 |
| DENSE†‡ [204] | $346 \times 260$ | RGB frames* | Depth maps, Segmentation | 8K samples | Depth, Semantic seg | - | 2020 |
| DSEC [130] | $640 \times 480$ | 2 cameras, RGB cameras, LiDAR, RTK GPS | Depth, Localisation | 53min | Depth | Switzerland | 2021 |
| DSEC-OF** [205] | | | Optical flow | | Optical flow | | 2021 |
| EventScape‡ [206] | $512 \times 256$ | RGB frames, Depth maps | Depth, Segmentation | 2h | Depth | Semantic seg | - | 2021 |
| Pedestrian-SARI§ [184] | $346 \times 260$ | RGB frames* | Detection | 141 sequences | "Person" | China | 2021 |
| DET [177] | $1280 \times 800$ | - | Lane extraction | 5h | - | Road lanes | China | 2019 |
| 1Mp Detection [170] | $1280 \times 720$ | - | Detection | 14h | "Car", "Pedestrian", "Two-wheeler" | France | 2020 |

* Available from the event camera itself.
† Extension of DDD17, providing ground truth to other problem.
‡ Extension of MVSEC, providing ground truth to other problem.
§ Not available for download, might be available upon request to the authors.
¶ Simulated data.
|| Distinct from DENSE dataset for LWIR and range-gated cameras [194] presented in Table I and in Section IV-B.
** Extension of DSEC, providing ground truth to other problem.
[32] M. P. Hansen and D. S. Malchow, “Overview of SWIR detectors, cameras, and applications,” in Proc. SPIE 6939, Thermosense XXX, 2008.

[33] F. Lv, Y. Zheng, B. Zhang, and F. Lu, “A Silicon Camera into an InGaAs Camera,” in IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[34] X. D. et al., “Timet: Object detection in thermal infrared images for autonomous driving,” Applied Intelligence, vol. 51, p. 1244–1261, 2020.

[35] S. Blundell and K. Blundell, Concepts in Thermal Physics. Oxford University Press, 2006.

[36] R. N. Hansen and D. S. Malchow, “Overview of SWIR detectors, arrays and their characterisations, techniques,” Defence Science Journal, vol. 59, pp. 580–589, 2009.

[37] J. Jung et al., “Infrared broadband metasurface absorber for reducing the thermal mass of a microbolometer,” Scientific Reports, vol. 7, no. 430, 2017.

[38] J. Tissot et al., “Uncooled microbolometer detector: recent developments at leti,” in Proc. SPIE 6940, Infrared Technology and Applications XXXIV, 2008.

[39] L. Yu et al., “Low-cost microbolometer type infrared detectors,” Micromachines, vol. 9, p. 800, 2020.

[40] F. Vicente, Z. Huang, X. Xiong, F. Torre, W. Zhang, and D. Levi, “Driver gaze tracking and eyes off the road detection system,” IEEE Transactions on Intelligent Transportation Systems, vol. 18, pp. 147–150, 2017.

[41] Y. Hong, Z. Hu, K. Uchimura, and M. Murayama, “Driver inattention monitoring system for intelligent vehicles: A review,” IEEE Transactions on Intelligent Transportation Systems, vol. 12, pp. 596–614, 2010.

[42] D. Dinges, “Sleepiness and the basis for alertness management,” 1998.

[43] M. J. Flores, J. M. Armigol, and A. de la Escalera, “Driver drowsiness detection system under infrared illumination for an intelligent vehicle,” IET Intelligent Transport Systems, vol. 5, no. 4, pp. 241–251, 2009.

[44] I. García, S. Bronte, L. M. Bergasa, J. Almazan, and J. Yebes, “Vision-based drowsiness detection for real driving conditions,” in IEEE Intelligent Vehicles Symposium, 2012.

[45] A. Dasgupta, D. Rahman, and A. Routray, “A smartphone-based drowsiness detection and warning system for automotive drivers,” IEEE Transactions on Intelligent Transportation Systems, vol. 20, no. 11, pp. 4045 – 4054, 2018.

[46] R. A. Naqvi, M. Arsalan, G. Batchuluun, H. S. Yoon, and K. R. Park, “Deep learning-based gaze detection system for automobile drivers using a night camera sensor,” Sensors, vol. 18, 2018.

[47] F. Xu et al., “Uncooled infrared microbolometer arrays and their characterisations, techniques,” Defence Science Journal, vol. 59, pp. 580–589, 2009.

[48] J. Jung et al., “Infrared broadband metasurface absorber for reducing the thermal mass of a microbolometer,” Scientific Reports, vol. 7, no. 430, 2017.

[49] J. Tissot et al., “Uncooled microbolometer detector: recent developments at leti,” in Proc. SPIE 6940, Infrared Technology and Applications XXXIV, 2008.

[50] L. Yu et al., “Low-cost microbolometer type infrared detectors,” Micromachines, vol. 9, p. 800, 2020.

[51] F. Niklaus et al., “Low-cost microbolometer type infrared detector array,” Sensors, vol. 12, no. 11, pp. 147–150, 2017.

[52] T. Tsuji, H. Hattori, M. Watanabe, and N. Nagaoka, “Development of an infrared camera sensor,” in Proc. SPIE 6836, MEMS/MOEMS Technologies and Applications III, 2007.

[53] R. Gade and T. Moeslund, “Thermal cameras and applications,” Machine Vision & Applications, vol. 25, pp. 245–262, 2014.

[54] J.-E. Kallhammer, “Night vision: requirements and possible roadmap for fire and air systems,” in Proc. SPIE 6198, Photonics in the Automobile II, 2006.

[55] O. Tsimhoni, J. Bargman, J. Minoda, and M. Flannagan, “Pedestrian detection with near and far infrared night vision enhancement,” 2004.

[56] O. Tsimhoni and M. Flannagan, “Pedestrian detection with night vision systems enhanced by automatic warnings,” in Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 2006.

[57] K. M. Judd, M. P. Thornton, and A. A. Richards, “Automotive sensing: A review,” Applied Intelligence, vol. 51, p. 1244–1261, 2020.

[58] J. Tissot et al., “Uncooled microbolometer detector: recent developments at leti,” in Proc. SPIE 6940, Infrared Technology and Applications XXXIV, 2008.

[59] N. Pinchon et al., “Overview of SWIR detectors, cameras, and applications,” in Proc. SPIE 6939, Thermosense XXX, 2008.

[60] F. Lv, Y. Zheng, B. Zhang, and F. Lu, “A Silicon Camera into an InGaAs Camera,” in IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[61] X. D. et al., “Timet: Object detection in thermal infrared images for autonomous driving,” Applied Intelligence, vol. 51, p. 1244–1261, 2020.

[62] S. Blundell and K. Blundell, Concepts in Thermal Physics. Oxford University Press, 2006.

[63] R. Bhan et al., “Uncooled infrared microbolometer arrays and their characterisations, techniques,” Defence Science Journal, vol. 59, pp. 580–589, 2009.

[64] J. Jung et al., “Infrared broadband metasurface absorber for reducing the thermal mass of a microbolometer,” Scientific Reports, vol. 7, no. 430, 2017.

[65] J. Tissot et al., “Uncooled microbolometer detector: recent developments at leti,” in Proc. SPIE 6940, Infrared Technology and Applications XXXIV, 2008.
K. Park, S. Kim, and K. Sohn, “Unified multi-spectral pedestrian detection based on probabilistic fusion networks,” Pattern Recognition, vol. 80, pp. 143–155, 2018.

G. Humbert-Renaux et al., “Thermal imaging on smart vehicles for person and road detection: Can a lazy approach work?” in IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), 2016.

K. Takumi et al., “Multispectral object detection for autonomous vehicles,” in Proceedings of the on Thematic Workshops of ACM Multimedia, 2017.

J. Wagner, V. Fisher, M. Herman, and S. Behnke, “Multispectral pedestrian detection using deep fusion convolutional neural networks,” in In Proceedings of 24th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN), 2016.

J. Liu et al., “Multispectral deep neural networks for pedestrian detection,” in Proceedings of the British Machine Vision Conference (BMVC), 2016.

C. Li et al., “Illumination-aware faster r-cnn for robust multispectral pedestrian detection,” Pattern Recognition, vol. 85, pp. 161–171, 2019.

— “Multispectral pedestrian detection via simultaneous detection and segmentation,” in In Proceeding of British Machine Vision Conference, 2018.

D. Guan et al., “Fusion of multispectral data through illumination-aware deep neural networks for pedestrian detection,” Information Fusion, vol. 50, pp. 148–157, 2019.

H. Zhang et al., “Multispectral fusion for object detection with cyclic fuse-and-refine blocks,” in IEEE International Conference on Image Processing (ICIP), 2020.

R. Yadav et al., “Cnn based color and thermal image fusion for object detection in automated driving,” in Irish Machine Vision and Image Processing (IMVIP 2020), 2020.

D. Konig et al., “Fully convolutional region proposal networks for multispectral person detection,” in IEEE Conference on Computer Vision and Pattern Recognition Workshop, 2017.

L. Zhang et al., “Weakly allighted cross-modal learning for multispectral pedestrian detection,” in IEEE/CVF International Conference on Computer Vision (ICCV), 2019.

K. Dasgupta et al., “Spatio-contextual deep network based multimodal pedestrian detection for autonomous driving,” arXiv:2105.12713, 2021.

G. Pelaez et al., “Road detection with thermal cameras through 3d information,” in IEEE Intelligent Vehicles Symposium, 2015.

J. S. Yoon et al., “Thermal-infrared based drivable region detection,” in IEEE Intelligent Vehicles Symposium, 2016.

L. Gillespie, “Apparent luminance as a function of range in gated, laser night-viewing systems,” Journal of the Optical Society of America, vol. 56, pp. 883–887, 1966.

I. M. Baker et al., “A low-noise laser-gated imaging system for long-range target identification,” in Proc. SPIE 5406, Infrared Technological and Applications, 2004.

A. M. Pinto and A. C. Matos, “Maresye: A hybrid imaging system for underwater robotic applications,” Information Fusion, vol. 55, pp. 16–29, 2020.

M. Bijelic et al., “Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020.

O. David, N. Kopeika, and B. Weizer, “Range gated active night vision system for automobiles,” Applied Optics, 2006.

N. Spooren et al., “RGB-NIR active gated imaging,” in Electro-Optical and Infrared Systems: Technology and Applications XII, 2016.

A. H. Williford et al., “Range-gated active short-wave infrared imaging for rain penetration,” Optical Engineering, vol. 60, no. 1, pp. 1 – 11, 2021.

F. Rutz et al., “Ingaas apd matrix sensors for swir gated viewing,” Advanced Optical Technologies, vol. 8, pp. 445–450, 2019.

S. Burri et al., “Architecture and applications of a high resolution gated spad image sensor,” Optics Express, vol. 22, pp. 17573–17589, 2014.

K. Morimoto et al., “Megapixel time-gated spad image sensor for 2d and 3d imaging applications,” Optica, vol. 7, pp. 346–354, 2020.

M. Bijelic, T. Gruber, and W. Ritter, “Benchmarking image sensors under adverse weather conditions for autonomous driving,” in IEEE Intelligent Vehicles Symposium (IV), 2018.

D. Konig et al., “Unification of range gated active imaging for 3d reconstruction,” in IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), 2020.
[138] G. Chen and A. Knoll, “Event-based neuromorphic vision for autonomous driving: A paradigm shift for bio-inspired visual sensing and perception,” IEEE Signal Processing Magazine, vol. 37, no. 4, pp. 34–49, 2020.

[139] T. Finate et al., “5.10 a 1280×720 back-illuminated stacked temporal contrast event-based vision sensor with 0.0665µm pixels, 1.066g/s mode programmable event-rate controller and compressive data-formatting pipeline,” in IEEE International Solid-State Circuits Conference, 2020.

[140] Y. Suh et al., “1280×960 dynamic vision sensor with a 4.95-µm pixel pitch and motion artifact minimization,” in IEEE International Symposium on Circuits and Systems, 2020.

[141] S. Chen and M. Guo, “Live demonstration: Cere-v: A 1n pixel multimode event-based sensor,” in IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019.

[142] Scheerlinck et al., “Ced: Color event camera dataset,” in IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019.

[143] N. Qiao, H. Mostafa, F. Corradi, M. Osswald, F. Stefanini, D. Sumialska, and G. Indiveri, “A reconfigurable on-line learning spiking neuromorphic processor comprising 256 neurons and 128k synapses,” Frontiers in Neuroscience, vol. 9, p. 141, 2015.

[144] F. Akopyan et al., “TrueNorth: Design and tool flow of a 65 mW 1 million neuron programmable neuromorphic chip,” IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 34, no. 10, pp. 1557–1557, 2015.

[145] M. Davies, A. Wild, G. Orchard, Y. Sandamirskaya, G. A. F. Guerra, P. Joshi, P. Plank, and S. R. Risbud, “Advancing neuromorphic computing with lohi: A survey of results and outlook,” Proceedings of the IEEE, vol. 109, no. 5, pp. 911–934, 2021.

[146] D. Gehrig, H. Rebecq, G. Gallego, and D. Scaramuzza, “EKLT: Asynchronous photometric feature tracking using events and frames,” International Journal of Computer Vision, vol. 128, no. 3, pp. 601–618, 2020.

[147] H. Akolkar, S. H. Ieng, and R. Benosman, “Real-time high speed motion prediction using fast aperture-robust event-driven visual flow,” IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1–1, 2020.

[148] U. M. Nunes and Y. Demiris, “Robust event-based vision model estimation by dispersion minimisation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1–1, 2021.

[149] M. Osswald, S.-H. Ieng, R. Benosman, and G. Indiveri, “A spiking neural network model of 3d perception for event-based neuromorphic stereo vision systems,” Scientific Reports, vol. 7, no. 1, p. 40703, 2017.

[150] T. Barbier, C. Teulliere, and J. Triesch, “Spike timing-based unsupervised learning of orientation, disparity, and motion representations in a spiking neural network,” in IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2021.

[151] A. Viale, A. Marchisio, M. Martina, G. Masera, and M. Shafique, “CarSNN: An efficient spiking neural network for event-based autonomous cars on the lohi neuromorphic research processor,” in International Joint Conference on Neural Networks, 2021.

[152] F. Paredes-Vallés, K. Y. W. Scheper, and G. C. H. E. d. Croon, “What can neuromorphic event-driven precise timing add to spike-based pattern recognition?” Neural Computation, vol. 27, no. 3, pp. 561–593, 2015.

[153] D. P. Moey et al., “Steering a predator robot using a mixed frame/event-driven convolutional neural network,” in Second International Conference on Event-based Control, Communication, and Signal Processing, 2016.

[154] ——, “PRED18: Dataset and further experiments with DAVIS event camera in predator-prey robot chasing,” arXiv:1807.03128, 2018.

[155] S. Afshar, T. J. Hamilton, J. Tapson, A. van Schaik, and G. Cohen, “Investigation of event-based surfaces for high-speed detection, unsupervised feature extraction, and object recognition,” Frontiers in Neuroscience, vol. 12, p. 1047, 2019.

[156] F. Paredes-Vallès, J. Hagenaaars, and G. de Croon, “Self-supervised learning of event-based optical flow with spiking neural networks,” arXiv:2106.01962, 2021.

[157] N. F. Y. Chen, “Pseudo-labels for supervised learning on dynamic vision sensor data, applied to object detection under ego-motion,” in IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2018.
using neuromorphic vision sensors;” in IEEE International Conference on Image Processing, 2020.

[184] J. Wan et al., “Event-based pedestrian detection using dynamic vision sensors,” Electronics, vol. 10, no. 8, p. 888, 2021.

[185] A. Mitrokhin, C. Ferrmüller, C. Parameshwara, and Y. Aloimonos, “Event-based moving object detection and tracking,” in IEEE/RSJ International Conference on Intelligent Robots and Systems, 2018.

[186] T. Stoffregen et al., “Event-based motion segmentation by motion compensation,” in IEEE/CVF International Conference on Computer Vision, 2019.

[187] Y. Zhou, G. Gallego, X. Lu, S. Liu, and S. Shen, “Event-based motion segmentation with spatio-temporal graph cuts,” IEEE Transactions on Neural Networks and Learning Systems, pp. 1–13, 2021.

[188] A. Mitrokhin, C. Fermüller, C. Parameshwara, and Y. Aloimonos, “Event-based moving object detection and tracking,” in IEEE/RSJ International Conference on Intelligent Robots and Systems, 2018.

[189] T. Stoffregen et al., “Event-based motion segmentation by motion compensation,” in IEEE/CVF International Conference on Computer Vision, 2019.

[190] Y. Zhou, G. Gallego, X. Lu, S. Liu, and S. Shen, “Event-based motion segmentation with spatio-temporal graph cuts,” IEEE Transactions on Neural Networks and Learning Systems, pp. 1–13, 2021.

[191] G. Chen et al., “EDDD: Event-based drowsiness driving detection through facial motion analysis with neuromorphic vision sensor,” IEEE Sensors Journal, vol. 20, no. 11, pp. 6170–6181, 2020.

[192] Y. Choi et al., “Kaist multi-spectral day/night data set for autonomous and assisted driving,” IEEE Transactions on Intelligent Transportation Systems, vol. 19, pp. 934–948, 2018.

[193] G. Choe et al., “Ranus:Rgb and nir urban scene dataset for deep scene parsing,” IEEE Robotics and Automation Letters, vol. 3, pp. 1808 – 1815, 2018.

[194] DENSE. (2019) Dense dataset. [Online]. Available: https://www. uni-ulm.de/en/in/driveu/projects/dense-datasets/

[195] Y. Hu, J. Binas, D. Neil, S.-C. Liu, and T. Delbruck, “DDD20 end-to-end event camera driving dataset: Fusing frames and events with deep learning for improved steering prediction,” in IEEE 23rd International Conference on Intelligent Transportation Systems, 2020.

[196] I. Alonso and A. C. Murillo, “EV-SegNet: Semantic segmentation for event-based cameras,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019.

[197] A. Sironi, M. Brambilla, N. Bourdis, X. Lagorce, and R. Benosman, “HATS: Histograms of averaged time surfaces for robust event-based object classification,” in IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018.

[198] A. Z. Zhu, L. Yuan, K. Chaney, and K. Daniilidis, “EV-FlowNet: Self-supervised optical flow estimation for event-based cameras,” Robotics: Science and Systems XIV, 2018.

[199] Y. Hu et al., “Slasher: Stadium racer car for event camera end-to-end learning autonomous driving experiments,” in IEEE International Conference on Artificial Intelligence Circuits and Systems, 2019.

[200] H. Rebecq, R. Ranftl, V. Koltun, and D. Scaramuzza, “High speed and high dynamic range video with an event camera,” IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1–1, 2019.

[201] S. Miao, G. Chen, X. Ying, Y. Yi, K. Ren, Z. Bing, and A. Knoll, “Neuromorphic vision datasets for pedestrian detection, action recognition, and fall detection,” Frontiers in Neurorobotics, vol. 13, 2019.

[202] P. de Tournemire, “A large scale event-based detection dataset for automotive,” arXiv:2001.08499, 2020.

[203] T. Fischer and M. Milford, “Event-based visual place recognition with ensembles of temporal windows,” IEEE Robotics and Automation Letters, vol. 5, no. 4, pp. 6924–6931, 2020.

[204] D. G. Javier Hidalgo-Carrio and D. Scaramuzza, “Learning monocular dense depth from events,” IEEE International Conference on 3D Vision, 2020.

[205] M. Gehrig, M. Millhäuser, D. Gehrig, and D. Scaramuzza, “E-raft: Dense optical flow from event cameras,” in International Conference on 3D Vision, 2021.

[206] D. Gehrig, M. Rüegg, M. Gehrig, J. Hidalgo-Carrio, and D. Scaramuzza, “Combining events and frames using recurrent asynchronous multimodal networks for monocular depth prediction,” IEEE Robotic and Automation Letters, 2021.

[207] T. Delbruck, R. Graca, and M. Paluch, “Feedback control of event cameras,” in IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2021.

[208] N. Khan, K. Iqbal, and M. G. Martini, “Time-aggregation-based lossless video encoding for neuromorphic vision sensor data,” IEEE Internet of Things Journal, vol. 8, no. 1, pp. 596–609, 2021.