Emergent Visual Sensors for Autonomous Vehicles

You Li, Julien Moreau, Javier Ibanez-Guzman

Emergent Visual Sensors for Autonomous Vehicles

You Li[1], Julien Moreau[2], Javier Ibanez-Guzman[1]

***Abstract*—For vehicles to navigate autonomously, they need to perceive and understand their immediate surroundings. Currently, cameras are the preferred sensors, due to their high performance and relatively low-cost compared with other sensors like LiDARs and Radars. However, their performance is limited by inherent imaging constraints, a standard RGB camera may perform poorly in extreme conditions, including low illumination, high contrast, bad weather (e.g. fog, rain, snow, etc.), glare, etc. Further, when using monocular cameras, it is more challenging to determine spatial distances than when using active range sensors such as LiDARs or Radars. Over the past years, novel image sensors, namely, infrared cameras, range-gated cameras, polarization cameras, and event cameras, have demonstrated strong potential. Some of them could be game-changers for future autonomous vehicles, they are the result of progress in sensor technology and the development of the accompanying perception algorithms. This paper presents in a systematic manner their principles, comparative advantages, data processing algorithms, and related applications. The purpose is to provide practitioners with an in-depth overview of novel sensing technologies that can contribute to the safe deployment of autonomous vehicles.**

I. Introduction

Since the dawn of the automotive industry, the dream of building autonomous cars never ends. The 2004 and 2007 DARPA Grand Challenges have demonstrated that vehicles could be driven unmanned in challenging conditions [3]. Since then, progresses in sensors, processors, and algorithms, have pushed autonomous vehicles to be considered strategic in industry and research. In parallel, to improve safety and comfort, advanced driver assistance systems (ADAS) such as *Lane Keeping Assistance* (LKA), *adaptive cruise control* (ACC), *automatic emergency braking* (AEB), are being integrated into modern vehicles. Currently, the automation level of vehicles is defined and classified by the SAE (*Society of Automotive Engineers*) [4]: From level 0 to level 3, the autonomous driving functions need the driver to be part of the control loop that they are usually named as ADAS. Level 4 and 5 allow for fully autonomous driving (AD) in restricted areas and anywhere. Under such context, the SIVALab [5], a joint research laboratory collaborated between Renault, UTC and CNRS, has been established in 2017 for the purpose of investigating the integrity and safety of autonomous vehicles.

Perception sensors, like human eyes, are critical for all levels of autonomous vehicles. Perception sensors include *visual sensors* like the cameras, and *range sensors* such as LiDARs [6], microwave and ultrasonic radars [7]. By using a *focal plane array* (FPA), a typical camera passively senses 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 standard video camera operates within the visible spectrum that each pixel value is represented by a combination of three basic colors, i.e. *red, green*, and *blue*. LiDARs and radars are sparse active range sensors that measure distance along 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 sensor data are then processed by a perception system to provide useful information for vehicle navigation. A typical perception system outputs two layers of information, as shown by C. Eising *et al.* [8]: 1) *Semantic* and 2) *Physical*. The semantic layer recognizes the objects of interest (e.g. pedestrians, vehicles, lane markings, traffic lights, etc), while the physical layer provides attributes to the classified objects: 3D positions, velocities, sizes, etc. In general, cameras are superior in the semantic layer, while LiDARs/radars are more reliable in locating objects. Accelerated by the breakthrough of deep neural networks (DNN), RGB cameras have been widely applied in all levels of autonomous vehicles as indispensable components in perception systems. For instance, usually combined with radars, monocular or binocular vision systems are used to detect pedestrians and vehicles for ACC and AEB functions, and detect lane markings for LKA functions. Nevertheless, Teslas autopilot system only utilizes several monocular cameras to realize ADAS without radars or LiDARs. To get rid of cost constraints, level 4 autonomous vehicles equipped with many cameras combined with multiple LiDARs and radars to create a 360 perception cacoon without blind zones. The applications of cameras in AD systems are similar to ADAS functions, but the applied scenarios are more complex and challenging.

Despite major successes, the limitations of RGB cameras in challenging situations have been recognized when used for safety-critical functions. Low illumination, glare, fog, rain, or other adverse conditions, can degrade their performance. For example, the glare generated by oncoming headlamps, and mirror-like reflections could blind the camera imager. Such image defects would lead to missed detections or unknown behaviors for a perception system, which might result in hazardous conditions. To enlarge ODDs (*operational design domain* [9]) and hence to improve safety, several emerging imaging technologies, e.g. the *infrared (IR) cameras, dynamic vision sensors (event cameras), polarization cameras, gated*

[1]You Li and Javier Ibanez-Guzman are with research department of Renault S.A.S, 1 Avenue du Golf, 78280 Guyancourt, France {you.li, javier.ibanez-guzman}@renault.com

[2]Julien Moreau is with Université de technologie de Compiègne, CNRS, Heudiasyc (Heuristics and Diagnosis of Complex Systems), CS 60319 - 60203 Compiègne Cedex, France julien.moreau@hds.utc.fr

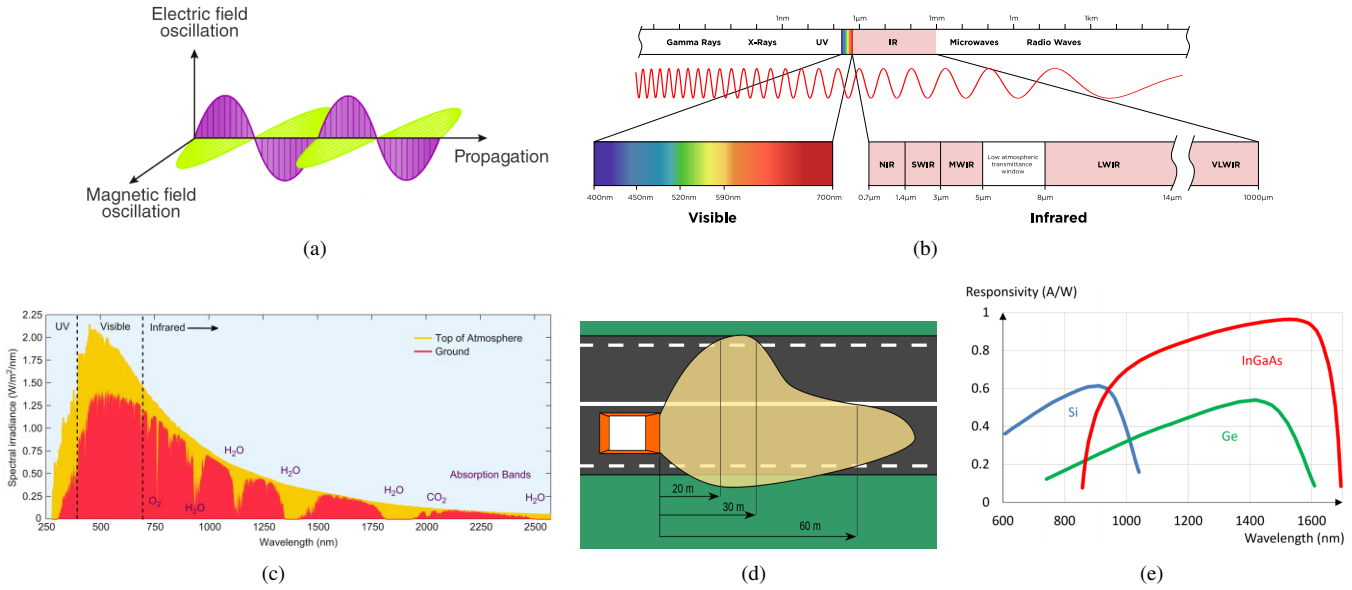


Fig. 1. (a) An example of electromagnetic (EM) waves. The mutually perpendicular electric field (in purple) and magnetic field (in purple) are periodically vibrating. (b) Electromagnetic spectrum. Visible light is a special kind of EM wave. (c) Solar irradiance spectrum as a function of wavelength (source from [1]). (d) At night, the area in front of the vehicle is illuminated by the low-beam headlamps. The maximum range is regulated to be around 60m. (e) Typical responsivity w.r.t wavelength for Si, InGaAs, and Ge-based image sensors (source from [2])

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.

In the literature, there are plenty of review papers on various aspects of autonomous vehicles. Yurtsever et. al. [10] give an overall introduction of all the modules in autonomous vehicles and envisaged future trends. Brummelen et. al. [11] focus on the current and future technologies for the perception of AD systems. The sensor fusion for both perception and localization is reviewed by Velasco-Hernandez el. al. [12] and Wang et.al. [13]. Marti el. al. [14] presents a review of the perception sensor technologies for automated driving systems. The included sensors are conventional such as visual spectrum (RGB) cameras, stereo cameras, millimeter-wave radars, LiDARs, and ultrasonic radars. Some reviews are around specific sensor technologies, such as the polarization camera [15], wide-angle camera [16], and the event camera [17]. Some focus on certain ADAS or AD functions, such as the perception of valet parking [18]. Nevertheless, other than the reviews on the conventional sensors and well-known perception systems, in this paper, we present a comprehensive survey of the emergent visual sensors that are designed to address the deficiency of the widely applied visual spectrum cameras. The introduced sensors and sensor data processing methods could broaden the horizons of the practitioners in ADAS/AD fields.

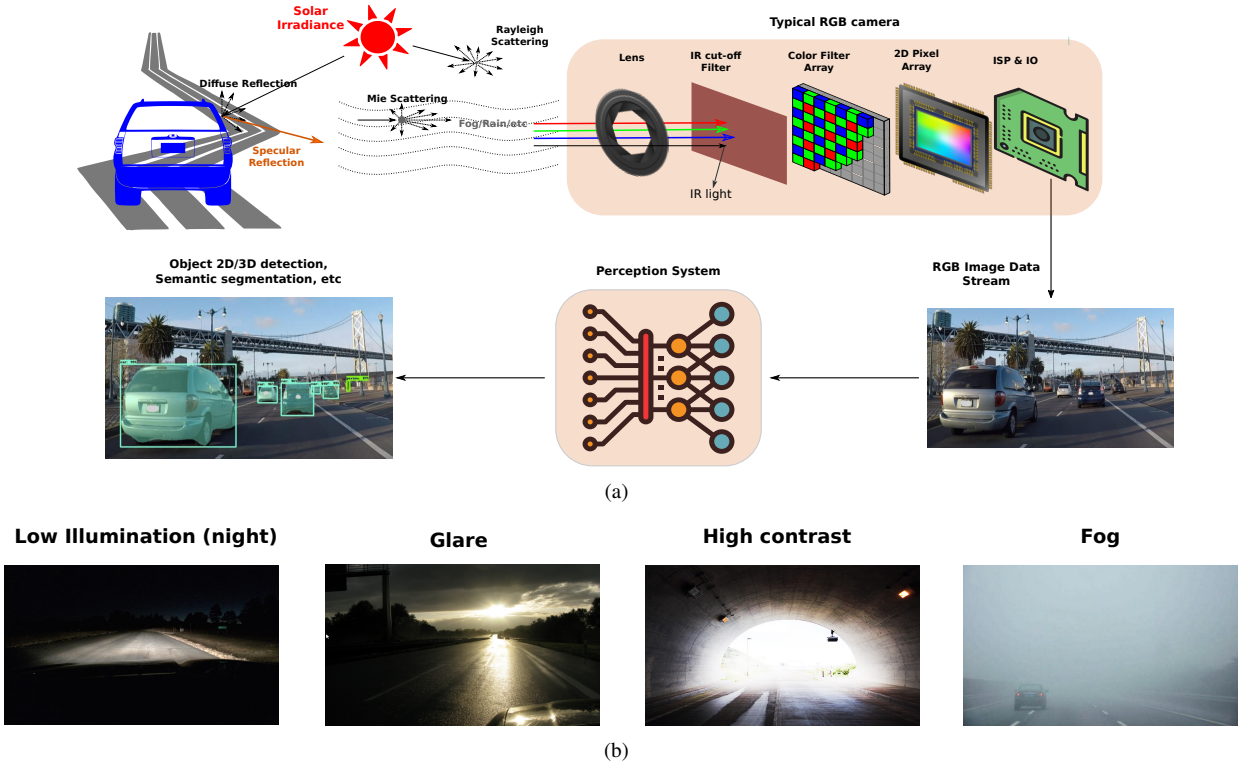
In this paper, a review of RGB cameras is firstly presented (Sec. II) to provide a baseline for comparative purposes. An analysis of the principles, applications, and associated algorithms for each of the selected cameras are included in the following sections: Section III for the infrared cameras, Section IV for the range-gated cameras, Section V for the polarization cameras and Section VI for the event cameras. Finally, Section VII concludes the paper and summarizes our findings as well as trends on the perceived technologies.

II. Principle of Conventional RGB Camera

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 λ of visible light ($\lambda \in [400nm, 700nm]$) is only a small portion of the EM spectrum ranging from Gamma rays ($\lambda < 1nm$) to radio waves ($\lambda > 1m$), 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 [19]). At night, vehicle headlamps and streetlamps are the primary sources of light [20]. 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 [21] (as shown in



(a)

(b)

Fig. 2. (a) A typical pipeline depicting how the light in a scene is converted to image pixels in an RGB camera before being processed by perception algorithms for environment understanding. The captured light from complex light reflections and scatterings is first passed the IRCF 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 leading to image over-saturation, and a foggy image caused by light scattering.

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 roads, 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 to or larger than a wavelength, such as the water droplet in fog, Mie scattering predominates [22].

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 an 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 semiconductor material, *Photodiodes* are the most commonly used components. A photodiode [23] 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 η represents the conversion efficiency of photons to electrons. For a specific wavelength λ, QE $\eta(\lambda)$ is defined as the percentage of photons hitting the photoreactive surface that produces EHPs:

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

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:

$$R(\lambda) = \frac{I_p}{P} = \eta(\lambda)\frac{q}{hf} \approx \eta(\lambda)\frac{\lambda}{1.24} \qquad [A/W] \qquad (2)$$

where q is electron charge, h is Planck's constant and f is the frequency of the optical signal. Fig. 1 (e) 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.

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 CFA, consisting of repeated 2×2 RGGB (red-green-green-blue) filter kernels because the human eye is more sensitive to green light. Only one of the three primary 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. RGB vector) for every pixel. Other types of CFA, such as RCCC or RCCB [24] (C stands for the wideband clear filter, i.e. no color filtering), are specifically designed for automotive applications. By only keeping red information for 1/4 pixels and not filtering light for all the other pixels, RCCC could improve the imaging sensitivity for both traffic signals and dark scenarios. However, full color is hard to be recovered by demosaicking algorithms. RCCB replaces the green pixels with clear ones to achieve low light sensitivity and hence lower noise. Unlike RCCC, the color information could be restored from RCCB as in [25]. As shown in Fig. 1 (e), 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.

C. Limitations and Latest Advancements

Conventional RGB cameras offer a good general purpose vision with a dense representation of the scene and its textures. Their limitations can be roughly classified as (1) *Image degradation in adverse conditions*, (2) *Motion blur in case of fast dynamics*, and (3) *Lack of 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 [28] 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 [29]. Fig. 2 (b) demonstrates such challenging scenarios for RGB cameras.

In the automotive industry, CMOS Image sensor (CIS) [30] [31] has dominated the markets because of lower cost and better imaging quality, compared with CCD image sensors. To accomplish the autonomous driving functions, the cameras are required to achieve high resolution for far object detection, wide dynamic range for high contrast lighting [32], high sensitivity for low light conditions [33], high frame rate for high-speed applications [34], and LED flickering mitigation [35] for stable traffic light recognition. Automotive CMOS cameras have made tremendous progress in the past decades. For instance, the image resolution and frame rate have reached 8MP pixels at 40fps for OmniVision's OX08B24C[1], or Sony's IMX324[2]. ONSemi's AR0821CS [36] has achieved a 150db dynamic range, which is very close to human eyes. For the nights, SONY's IMX390 [37] can stable output colorful images at 0.1lux, which is equivalent to moonlight.

In parallel, computer vision algorithms, especially the deep neural network (DNN) based methods, gained unbelievable progress in the past ten years that the commercial ADAS prevail in current passenger vehicles and various L4 autonomous driving start-ups appeared across the world [38]. Among various DNN structures, convolutional neural networks (CNNs) show superior performance in computer vision because the convolutional operations can efficiently capture spatial features from images. In 2015, CNN-based image classification in ImageNet surpassed human performance for the first time in history [39]. In ADAS/AD, the CNNs are principally used for object detection, with classic algorithms such as YOLO [40], [41], fasterRCNN [42] and SSD [43], etc. However, in recent years, transformers [44] have surpassed CNNs and they almost dominated state-of-the-art performance in most computer vision fields. For instance, object detection [45], lane marking detection [46], and sensor fusion [47]. Its believed that in the future, more and more powerful algorithms will be coming to squeeze the information of each pixel.

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 humans and is often divided as follows: (1), *Near-infrared* (NIR): wavelength ranging from 0.7μm to 1.4μm. (2), *Short-wavelength infrared* (SWIR): wavelength ranging from 1.4μm to 3μm. (3), *Long-wavelength infrared* (LWIR, or Far-infrared (FIR)): wavelength ranging from 8μm to 14μm. 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 [48]. 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 similarly to RGB imagers in that they directly transform photons into electrical signals. While

[1] https://www.ovt.com/sensors/OX08B4C
[2] https://www.sony-semicon.co.jp/products/common/pdf/IMX324_424.pdf

(a) (b)

Fig. 3. (a) A RGB-IR imaging system: the CFA is replaced by an RGB-IR filter array to capture RGB and infrared intensities (from [26]) (b) RGB (left) and NIR (right) images for the same scene. Vegetation in the NIR spectrum is "brighter" than in RGB image (from [27]).

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.

A. NIR camera

NIR imagery shares many properties with RGB imagery: as shown in Fig. 1 (e), a silicon-based imager can still exhibit NIR sensitivity until around 1100nm. As a result, with proper modifications, an RGB camera can be converted into a NIR camera. Because the CFA still has transmission spectra that bleed into NIR wavelengths, removing the IRCF or replacing it with a NIR bandpass filter converts a consumer-grade RGB camera to a NIR camera, as demonstrated in [49] and [50]. Fig. 3 (b) shows an 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 [51] achieves 50% QE at 940nm, and 70% QE at 850nm NIR wavelength. Although other types of materials, such as InGaAs [52] 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.* [53] present a four-bandpass filter array to acquire RGB-NIR images. In Lu *et al.*'s work [54], a 4×4 pattern containing 15 visible/NIR filters and 1 NIR-only filter, is made. Park *et al.* [55] and Skorka *et al.* [56] 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 [26] for automotive applications, as shown in Fig. 3 (a). A more comprehensive study on RGB-IR camera design could be found in Geelen *et al.* [57].

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 [58], e.g. structured light-based 3D reconstruction. Two popular wavelengths are 850nm and 940nm. In the early days, 850nm NIR emitters were used because of their 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 [59] as well.

Though silicon-based image sensors have excellent responsivity 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 [60], as shown in Fig. 1. In comparison to other semiconductor materials used in the SWIR spectrum e.g. Ge or HgCdTe (*Mercury Cadmium Telluride*), InGaAs detectors are cost-effective and high-sensitive while being operated at room temperature [61]. However, compared to silicon-based sensors, InGaAs detectors suffer issues of the higher fabrication cost and pixel detects. Here, we mainly review the characteristics and applications of NIR and LWIR cameras applied in the automotive industry.

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 λ is a function of temperature given by Planck's law [63]:

$$B(\lambda, T) = \frac{2hc^2}{\lambda^5} \frac{1}{e^{hc/\lambda k_B T} - 1} \tag{3}$$

(a) (b) (c)

Fig. 4. (a) The architecture of a microbolometer. When exposed to LWIR radiance, the absorber generates heat, which is measured by a thermometer and converted into electrical signals. (b) and (c) RGB images (left) and thermal images (right) of the same scenes (from [62]). The RGB image quality is severely degraded by low illumination, glare, and fog, whereas the thermal camera is unaffected and provides clear images for object detection.

where λ, h, c, K_B are the wavelength, Planck's constant, light speed, and Boltzmann's constant. The standard unit of $B(\lambda, T)$ is $W \cdot sr^{-1} \cdot m^{-3}$. Most of the radiation emitted by the human body is mainly at the wavelength of 12μm, which is located in the LWIR spectrum. That's the reason for using an 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* [64] is an instrument that measures heat radiation and converts it into certain measurable quantities. Fig. 4 (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 [65]. 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.* [66], Yon *et al.* [67]. The thermal measures are then transferred to the ROIC for further processing.

A 2D microbolometer array [68] can capture thermal images with a much more affordable price and compact size that it is particularly well-suited for mobile applications such as the automobile. 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. For the uncooled microbolometer imaging systems, pixel non-uniformity and temperature fluctuations from the ambient conditions result in thermal drift and spatial nonuniformity of each pixel. To overcome those disturbing influences for better image quality, a series of compensation procedures such as non-uniformity correction (NUC, or referred as a flat-field correction) [69], gain correction, offset correction, and radiometric calibration are carried out. In traditional thermal vision systems, an optical shutter is utilized for a run-time re-calibration during its closure time. However, such a shutter-based mechanism would interrupt the imaging processing and impede more compact sensor sizes. For automotive applications, shutter-less thermal vision systems [70] [71] being capable of self-calibration have more advantages. From the industry side, commercially available shutterless thermal cameras are off-the-shelf, such as the BOSON camera from Teledyne [72], Adasky camera [73], and Lynred camera [74].

D. Advantages

The principal advantages of IR cameras are the capabilities of addressing adverse weather conditions and low illumination conditions. Operating without the need of external light sources, 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. Furthermore, the LWIR imagery does not suffer from the glare effects caused by the facing headlamps [75]. Fig. 4 (b) - (c) show a comparison of thermal imagery and visible imagery in several harsh conditions. Coupled with one or more invisible NIR transmitters, an active NIR camera could be a cost-effective substitute for a thermal camera. Because an NIR camera behaves similarly to the visible spectrum as "reflective infrared", the NIR imagery provides more contextual information, e.g. driver's lane markings, texts in traffic signs, etc. Such information enables NIR cameras to offer more functions (e.g. lane-keeping assistance, driver monitoring systems) at night.

E. Applications in Autonomous Vehicles

Automotive night vision system (NVS) is one of the key areas utilizing NIR or LWIR cameras. In 2000, General Motors launched the first automotive NVS on the Cadillac DeVille using an LWIR sensor supplied by Raytheon [79]. In 2004, Honda [80] introduced a thermal camera-based Intelligent NVS on Honda Legend. In 2005, BMW began to use LWIR cameras in its 7 Series. Peugeot incorporated a thermal camera into its flagship sedan Peugeot 508 in 2018. In 2002, Toyota presented an active NIR NVS in Toyota LandCruiser and Lexus 470, but in 2014 Lexus decided to discard the NVS in the subsequent generations. In 2022, Kyocera [81] announced a night vision system consisting of a vehicle headlight that can emit both white and NIR light on the same optical axis, and a vehicle-mounted RGB-NIR camera to detect objects.

Many comparisons and discussions have taken place between the active NIR cameras and passive LWIR night vision systems as in Kallhammer [82]. In general, it has been

(a) (b) (c)

Fig. 5. (a) An example of a DMS consisting of a camera and an IR illuminator facing the driver to detect whether the driver's eyes are on the road or not (from [76]). (b) Face landmark detection results (from [77]). (c) 15 zones for gaze classification (from [78]).

demonstrated in Tsimhoni *et al.* [83] and [84] 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.* [85] show that LWIR imaging is significantly less affected by fog than NIR cameras. The experiments conducted by Pinchon *et al.* [75] 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. A recent evaluation (Velazquez *el al.* [73]) of thermal image based pedestrian detection under fog shows that, when the fog's visibility is less than 20m, a thermal camera (VGA resolution, 60HFoV) can still reliably detect a pedestrian at 25m.

On the other hand, the tests in Pinchon *et al.* [75] show that thermal cameras are unable to detect lane markings or recognize traffic signs, whereas NIR imaging systems can. Keeping contexture information let NIR cameras dominate the market of driver monitoring systems (DMS). For instance, in the DMS named SuperCruised launched by Cadillac in 2018, one NIR camera is mounted in the instrument panel to monitor whether or not the driver is watching the road.

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.A. involve a certain form of driver inattention [86]. *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 keeps attention on the road, as shown in Fig. 5 (a). A typical DMS usually contains *gaze detection* and *drowsiness detection* to warn the driver when an inattention event is detected. Researches (Ahlstrom *et al.* [87], Schwarz *et al.* [88]) have proved that the DMS could effectively improve safety. In Europe, a general safety regulation[3] has been passed in 2019 to mandate automakers to install advanced safety systems including DMS in new cars in 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. 5 (b) shows an example of detected facial landmarks. In recent years, DNNs dominate this domain. For example, Yoon *et al.* [78] utilize a VGG network for face detection and Park *et al.* [89] 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.* [90] is a valid metric for detecting drowsiness that has been used in many studies, e.g. Ji *et al.* [91], Flores *et al.* [92], Garcia *et al.* [93] and Dasgupta *et al.* [94]. 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 on the 3D gaze estimation via 3D modeling of face/eyes. As in the AttenD algorithm proposed by Ahlstrom *et al.* [87], the estimated 3D gaze direction is compared with a predefined 3D safe region to detect distraction events. Vicente *et al.* [76] 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. 5 (c) shows an example of 15 divided gaze zones for gaze classification. Fridman *et al.* [77] and Naqv *et al.* [95] 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.* [78] 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.* [96] and Akinyelu *et al.* [97].

2) LWIR cameras in Night Vision Systems: Before the era of deep learning, object detection followed a traditional pipeline as: candidate region proposal, feature extraction and machine learning based classification, such as Haar feature-based cascade AdaBoost classifier [98], SVM classifier [99]. Fang *et al.* [100] manually design features from hotspots in a thermal image to train a SVM classifier to recognize

[3]https://ec.europa.eu/commission/presscorner/detail/en/IP_19_1793

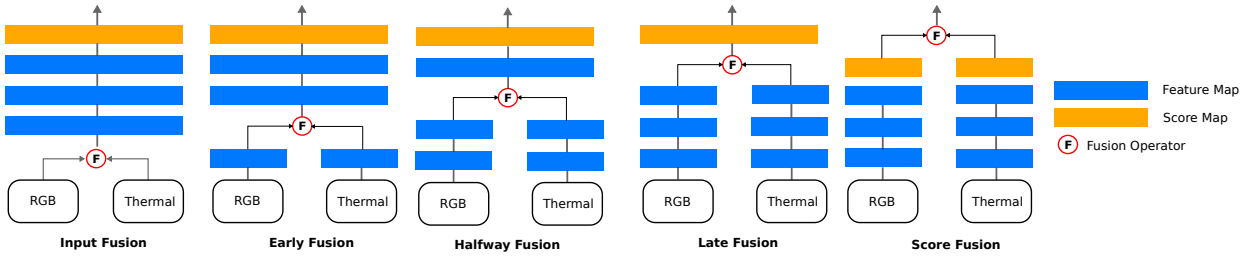


Fig. 6. Different strategies of fusing RGB and thermal images according to specific stages.

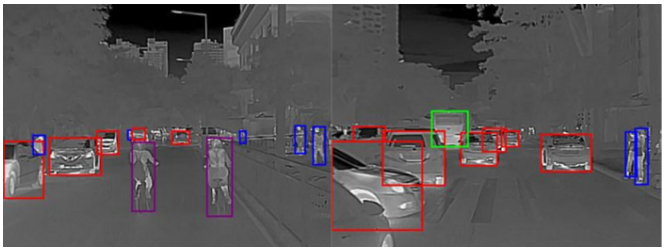


Fig. 7. Object detection in thermal images. Pedestrians, vehicles, etc are localized in 2D bounding boxes with different colors. (from [62]).

pedestrians. Forslund *et al.* [101] present a large animal thermal image dataset gathered over an 8-year driving period and a cascade AdaBoost classifier for animal detection.

Entering into the deep learning era when the CNNs sweep all the computer vision benchmarks, there is no exception in thermal image processing. Kristo *et al.* [102] benchmark several popular object detectors, including Faster R-CNN [42], SSD [43] and YOLOv3 [103], 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.* [62] propose a TIRNet for pedestrian detection (as shown in Fig. 7) by modifying the SSD detector. The performance of TIRNet is reported better than YOLOv3 based on their annotated dataset and the KAIST dataset [104]. A large-scale thermal pedestrian dataset SCUT is presented in Xu *et al.* [105]. Based on this dataset, the authors provide a detailed comparison between widely used detectors. Tumas *et al.* [106] present a ZUT dataset containing vehicle odometry and weather measures. Launched in 2019, an EU project HELIAUS [4] aims to promote a thermal perception system for both in-cabin passengers monitoring and exterior object detection for ADAS in all light conditions. Within this project, Farooq *et al.* [107] [74] elaborately evaluate the performance of a YOLOv5 [108] detector based on a shutterless thermal camera. Meanwhile, an annotated thermal automotive dataset C3I-ADAS [109] (36K 640×480 thermal images with annotations) is released under the HELIAUS project for research purposes.

F. Fusion of multiple spectrum images

Infrared images and visual spectrum images are complementary to each other. Fusion of the images from multiple spectrum enable a perception system have robust performance both in daytime and at night, as well as harsh weathers. Due to the length limitations, we take the fusion between LWIR and visual spectrum as an example to introduce the principles and mainstream methods of fusion.

1) Fusion in CNN framework: fusion between thermal and RGB images is still indispensable in practice to address perception issues in all lighting conditions [110]. Under the framework of CNN, the fusion can be performed in various stages within a typical CNN architecture, and accordingly be roughly divided as *input fusion*, *early fusion*, *halfway fusion*, *late fusion* and *score fusion*, as illustrated in Fig. 6. Late fusion offers the flexibility to directly fuse existing detectors inferring in parallel. Choi *et al.* [111] and Park *et al.* [112] fuse this way two CNNs for proposal generation on color and thermal streams. With more modalities, Humblot-Renaux *et al.* [113] investigate the late fusion for multispectral people detection from YOLO detectors, as well as Takumi *et al.* [114] from RGB, NIR, MIR and LWIR images. Other authors, e.g. Wagner *et al.* [115], Liu *et al.* [116], Li *et al.* [117], 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.* [118], Guan *et al.* [119] and Yadav *et al.* [120].

Another trend is looking for new neural network modules as fusion operator. Zhang *et al.* [121] propose a "Cyclic Fuse-and-Refine" module to optimize the complementary and consistency of multispectral features. In Li *et al.* [117] and Guan *et al.* [119], illumination detection modules are proposed to dynamically assign the weights of multispectral features under a halfway fusion architecture. Dasgupta *et al.* [122] extend the halfway fusion architecture with a multimodal feature embedding module (MuFEm) and a CRF-based Spatial-Contextual feature aggregation module.

2) Fusion based on attention model and Transformer: The above fusion methods within CNN architecture are restrained by the CNNs' limited expressive capability. The convolution operator is regarded as a non-fully connected graph with a local receptive field that only integrates local information. In contrast, an attention operator acts as a fully connected graph to have a global receptive field to learn long-range dependencies for more complex feature expressiveness. In recent years, attention models, in particular, the visual transformers are regarded to be more suitable in multi-modality fusion than tradi-

[4] https://www.heliaus.eu/news-events/

tional CNN methods. A well-designed "cross-modality fusion" (CMF) transformer proposed in [123] is used to s used to fuse the intra-modality and inter-modality features simultaneously. The experiments show that, after the enhancement of CMF, a modified YOLOv5 detector demonstrates better performance than the original version. Shared with similar ideas, Y. Zhang *et al.* [124] propose a channel-wise attention module (CAM) and a spatial-wise attention module (SAM) for multi-spectral feature fusion applied in pedestrian detection. Q. Fang *et al.* [125] propose a cross-modality feature complementary module working on channel-wise feature fusion and an attention-based feature enhancement fusion module working on spatial feature aggregation before the detection head. K. Dasgupta *et al.* [126] use a graph attention module to extract multi-spectrum features and a feature fusion unit to address the modality imbalance problem. In general, the proposed attention models are mainly used for enhanced feature extraction and fusion, the fused features are then sent to certain typical detection heads for final object detection. Some popular open-sourced multi-spectrum datasets are listed in Tab. II.

G. Remaining Challenges

Both the NIR and LWIR cameras have been successfully applied in mass-produced cars for night vision systems or driver monitoring systems. For LWIR cameras, the main challenge is to increase the resolution while still maintaining affordable prices. The ordinary resolution for an LWIR camera is still VGA (640×480 = 0.3M pixels), which is much smaller than a standard RGB camera used for ADAS/AD (2M pixels to 8M pixels). Another issue is the integration position. Because regular glass is opaque to thermal radiation, a thermal camera has to be mounted outside the windshield, typically placed in the front grille. Such a position makes the thermal camera vulnerable to being damaged by road debris or covered by snow/ice/dust. As for NIR cameras, they are more mature in production due to the shared fabrication technology with visible spectrum imaging sensors. However, to be performance comparable with LWIR camera at night, more advanced algorithms or hardware improvements are expected to increase NIR's imaging stability in challenging lighting conditions, such as the oncoming vehicle's headlamp at night. According to the experiments in Boullough [20], the spectrum of an ordinary halogen headlamp contains a large portion of NIR wavelengths ranging from 700nm to 1000nm. Those emitted NIR lights would cause image clutters and thus reduce the performance of a perception system.

IV. Range-Gated Camera

To enhance the imaging quality under harsh conditions, range-gated imaging was first proposed in the 1960s [127] and has been applied in night vision systems [128], submarine vision [129]. In recent years, range-gated cameras have gained popularity for their resistance to adverse conditions [130].

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 main 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.* [131] and Spooren *et al.* [132]. When considering better penetration in long-distance through fog or smoke, the SWIR laser is preferred because it can achieve much higher transmission power while still meeting eye safety standards. In [128], a range-gated imaging system based on an Nd YAG laser at 1571nm reaches a 10km detection range. Similarly, in Baker *et al.* [133], 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.* [132], a gated RGB-NIR image sensor with high NIR quantum efficiency (≈40%) is built. In Rutz *et al.* [134], 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.* [135] present a 512×128 pixel CMOS SPAD sensor capable of operating within an exposure window as small as 4ns. In Morimoto *et al.* [136], a 1M pixel CMOS SPAD image sensor is built for 3.8ns gating time.

B. Advantages

The range-gated cameras, as active sensors, are better suited to low-light conditions, such as a country road at night. Furthermore, owing to the slicing 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.

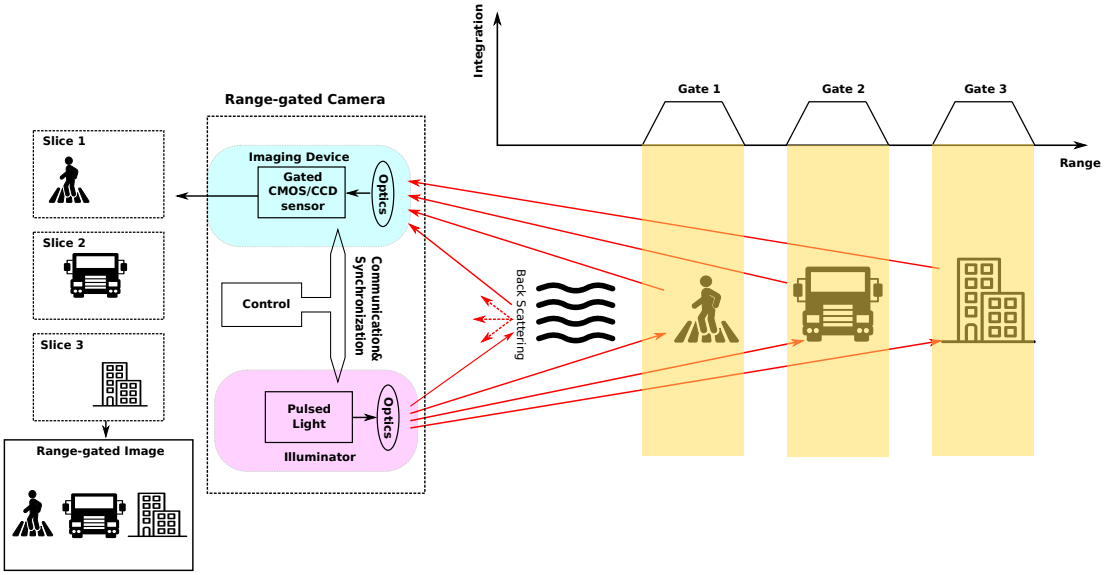


Fig. 8. The imaging principle of a range-gated camera: merging several image slices into a final image. A very short exposure defined by a gating function yields an image slice. The backscattering interference outside the range defined by a gating function has almost no effect on the outcome.

According to [131], a key parameter deciding the quality of a range-gated image is the modulation contrast:

$$Contrast \simeq \frac{I_{target} - I_{background}}{I_{target} + I_{background} + 2I_{bsc}} \tag{4}$$

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

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

where R_{on}, R_{off} define a range interval during one exposure. G is the backscatter gain, γ is the atmospheric attenuation coefficient, θ is the laser beam divergence, P is the laser power, F_n is the speed of the lens, and X is the integration variable. Compared with a conventional camera counting all the backscattering photons, a range-gated camera only performs the photon integration during a very short opening time, i.e. between R_{on} and R_{off}, so that a higher contrast defined in Eq. 4 can be achieved.

C. Applications in Autonomous Vehicles

Owing to its excellent performance in harsh conditions, the range-gated camera has the potential to be a strong competitor to infrared cameras, and has gained recognition in recent years. A comparison between an RGB camera and a range-gated camera in a fog environment is shown in Fig. 9. Walz *et al.* [138] benchmark multi-model sensors in a well-controlled artificial fog chamber. Both the quantitative and qualitative results show the superiority of range-gated cameras. On the industry side, [139] first applies a NIR range-gated camera to aid driving at night. Grauer *et al.* [140] and [141] 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 were developed within the EU-founded DENSE project [5]. Supported by this project, the DENSE dataset [6] containing multi-model sensors (a range-gated camera, an RGB stereo camera, an LWIR camera, and a LiDAR) is released to the public. The dataset covers snow, rain, and urban and suburban scenarios. The DENSE dataset is further annotated in Julca-Aguilar *et al.* [142] 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.* [143] present a deep neural network (DNN) named "*gated2depth*", which can estimate the depth of each pixel in the range-gated camera. The proposed DNN architecture utilizes all three slice images. Walz *et al.* [138] extend *gated2depth* by incorporating aleatoric uncertainties into the pixel-wise depth estimation. Bijelic *et al.* [130] propose a fusion neural network 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.* [142] propose a DNN for 3D object detection. The proposed DNN is tailored to the temporal illumination cues from the three image slices. Based on the *Gated3D* dataset, they demonstrated that using temporal cures from a range-gated camera, the 3D object detection results outperform a pure RGB-based detection method.

D. Remaining Challenges

As an active sensor, a range-gated camera needs an illuminator precisely synchronized with the image sensor which

[5] https://www.dense247.eu/
[6] https://www.uni-ulm.de/en/in/driveu/projects/dense-datasets/

Fig. 9. Imaging results of a conventional RGB camera (left) and a range-gated camera (right) in an artificial fog (from [137]). Due to the backscattering effect caused by the fog, many objects are obscured in the RGB camera image, while the range-gated camera image is almost immune to the fog. Meanwhile, the headlamp of the target car causes a strong blooming effect in the RGB camera image but has no impact on the range-gated camera image.

is more complicated than passive cameras in practice. The illuminator may potentially cause interferences with other sensors such as visual spectrum cameras or LiDARs. In addition, because they are novel, range-gated cameras are rare in usage compared to other cameras. Therefore, there is still a lack of large open-sourced datasets to boost relevant studies to explore the capabilities of the range-gated camera.

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 [146]. 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 [147], 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 $\mathbf{S}$ [148], consisting of 4 parameters: $\mathbf{S} = [S_0, S_1, S_2, S_3]$. $S_0 (> 0)$ 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.

$$\begin{aligned} S_0 &= I_0 + I_{90} = I_{45} + I_{135} \\ S_1 &= I_0 - I_{90}, \quad S_2 = I_{45} - I_{135} \end{aligned} \tag{6}$$

where I_0, I_{45}, I_{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:

$$AoP = \frac{1}{2} \times arctan(\frac{S_2}{S_1}), \quad DoP = \frac{\sqrt{S_1^2 + S_2^2}}{S_0} \tag{7}$$

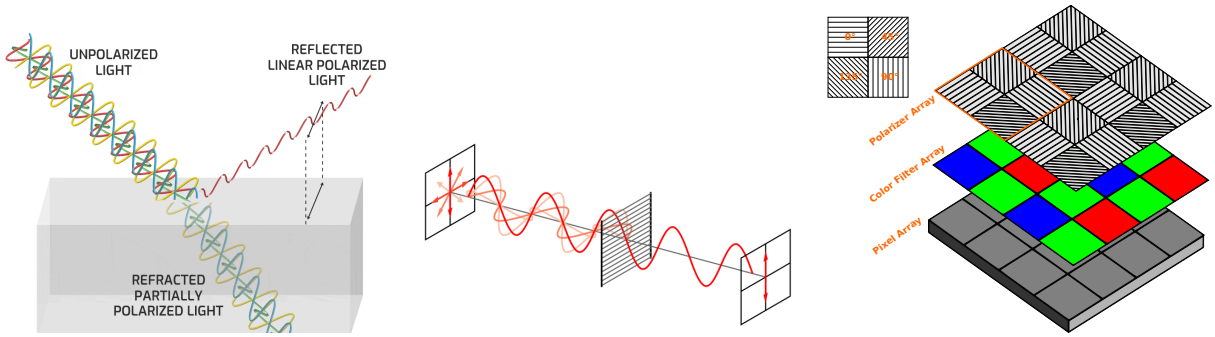
Varying between 0 and 180, AoP represents the predominant axis of the light vibration. DoP is the ratio of the intensity of the polarized portion to the total intensity. For instance, linearly polarized light has a DoP of 1, and natural light usually has a DoP between 0 to 0.5.

Creating a practical and convenient polarimetric imaging system is not easy work. In early research, Morel *et al.* [149] make 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.* [150], 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 has been employed in many computer vision studies.

B. Advantages

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.* [151], and Wang *et al.* [152]) could alleviate these issues, polarization cameras offer a new perspective. One of the biggest advantages is that the polarization camera can detect transparent objects like windows or glasses [153], which are hard to be detected by ordinary cameras or LiDARs. The other man-made objects, like the vehicles, are more distinct in polarization images [154], [155].



(a) (b) (c)

Fig. 10. (a) Unpolarized light beams could be converted to polarized light after reflection. (from [144]). (b) A polarizer's function is to convert an unpolarized beam into a (linear) polarized beam (from [145]). (c) A schematic diagram of SONY IMX250 CMOS sensor. To acquire color and polarization information, a micro-polarizer array and a color filter array are placed on top of pixels.

C. Applications in Autonomous Vehicles

Polarization cameras are not yet commercialized for automotive usage. 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*. Wang *et al.* [156] 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.* [157], 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.* [154] 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 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.* [160] and [161] show that car detection results under adverse weather conditions could be improved by 20% - 50%. In addition, a new dataset PolarLITIS [162] 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.* [163] and Xiang *et al.* [164] both find that the semantic segmentation for car and windows is largely improved thanks to the polarization features.

D. Remaining Challenges

Polarization measures as a new data dimension would benefit the perception system. While, since a polarization camera needs a micro-polarizer array, it sacrifices its imaging performance in dark conditions, which is a critical issue for autonomous vehicles. Another problem is similar to the range-gated camera, polarization cameras are pretty young for researchers in robotics/autonomous vehicles fields, and more open-sourced datasets and related studies are needed to show the unique advantages brought by polarization.

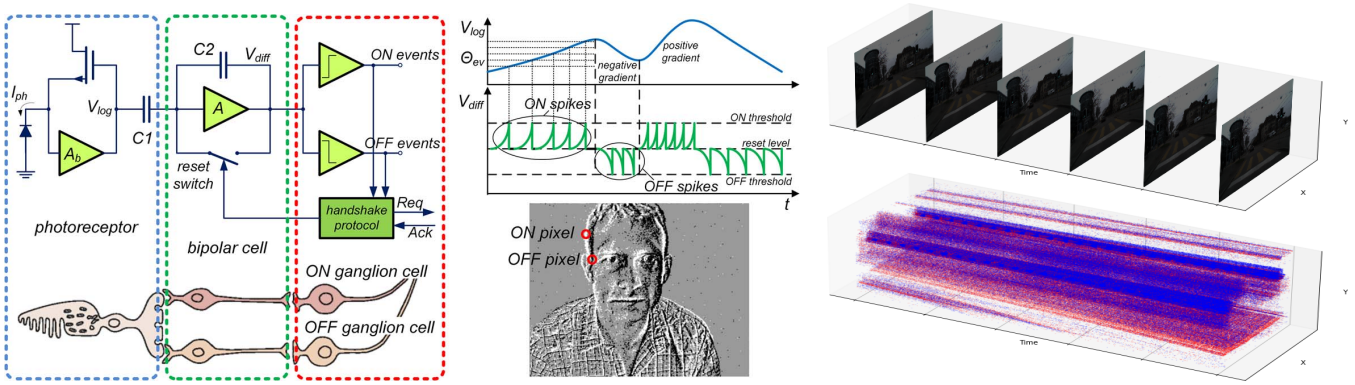
VI. Event Camera

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 [165], and a tutorial aiming at some common processing methods applied for autonomous driving is given in [166]. 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 [158]. In the retina, the fundus of the eye, are located the cones and rods, which 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



(a) (b)

Fig. 11. (a) Retinomorphic event vision with spiking output (from [158]). Human retina layers and corresponding dynamic vision sensor circuit for one pixel (left). Sample of the signal waveform (top right), and the response of an array of such pixels accumulated for a short time, for a sample scene showing a moving person (bottom right). (b) RGB frames from a standard camera (on top), compared to events flow (bottom), from the same time-synced scene issued from DSEC dataset [159]. Red points depict positive events, while blue points depict negative events.

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). Similarly 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 filters 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) [167], Samsung DVS-Gen4 (1280×960 pixels) [168], CelePixel CeleX-V (1280×800 pixels) [169]. Some event cameras (e.g. iniVation DAVIS346[7], CelePixel CeleX-V) incorporate 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[8], 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) [170].

[7]http://inivation.com/wp-content/uploads/2020/09/DAVIS346.pdf

[8]http://inivation.github.io/inivation-docs/Hardwareuserguides/User_guide_-_DAVIS_USB3_development_kit.html

B. Advantages

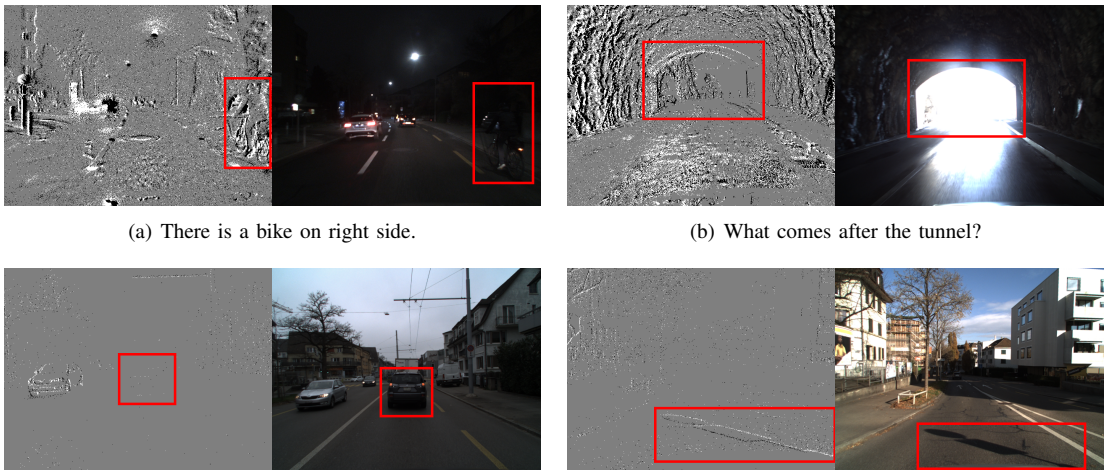
Event cameras are bio-inspired passive sensors that try to imitate millions of years of evolution of sight sense. General advantages of event camera are stated by Gallego *et al.* [165]: *Microsecond temporal resolution* for detection and timestamp. A direct consequence is the ability to always avoid motion blur as it exists for conventional cameras. Furthermore, the event camera outputs at sub-millisecond latency, which is approximately equivalent to a virtual $> 1000FPS$ 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 to detect fast-moving entities; 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 Fig. 12 (a), (b).

C. Data representation and processing

Unlike frame cameras, neuromorphic cameras output stream of events $\mathcal{E} = \{e \forall (x, y, t, p)\}$ that each event $e(x, y, t, p)$ encodes pixel 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 $e = p \cdot \delta(\xi - x, \upsilon - y) \cdot \delta(\tau - t)$ 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 arises, 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



(a) There is a bike on right side.

(b) What comes after the tunnel?

(c) Front car's speed equals ours so it is not visible in events.

(d) With events, shadows can look like noise or even obstacles.

Fig. 12. Advantages (a), (b), and limits (c), (d), of event camera images compared to normal images: (a) Night scene, (b) Glare light, (c) Stationary car, (d) Shadow on the ground. Scenes taken from DSEC dataset [159].

current computers are not designed for spikes processing. Standard processor 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 [171], IBM *TrueNorth* chip [172], and Intel *Loihi* chip [173]. 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 [174]–[176]), most of them integrate asynchronous events in artificial neural networks. SNNs have already been proposed for many applications. For example, [177], [178] for stereo depth estimation, [179] for classification, [180] for optical flow, [181] for background motion separation, [182], [183] for heading estimation and loop closure detection, [184], [185] for robotic control, [186] for target following, [187] 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, allowing for example Brebion *et al.* [188] to optimize a pipeline for real-time optical flow from HD events stream. 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 but 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.* [189] use adaptive accumulation time (making the method asynchronous), Rebecq *et al.* [190] use overlapping pairs of time windows, and Joubert *et al.* [191] combine time windows of different lengths. The study [192] 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 [193]–[196]. It allows for keeping accumulated representations with similar appearances (same density of events) and for asynchronous processes. Nevertheless, the following operations should be fast enough when huge flows of events arrive in a short time. At last, event accumulation can be motion compensated with a fast algorithm, typically using a joint IMU inside the event camera [190]. This guarantees event binning with no blur effect in case of a long accumulation time.

After accumulating enough events, the next issue is *encoding*, i.e. extracting effective event attributes. Several *hand-crafted encoding* methods have been proposed in the literature (for interested readers, a categorisation is suggested in [197]). For example, leveraging "frequency encoding" representation from Chen *et al.* [198], where a standard YOLOv3 CNN architecture [103] is utilized for pedestrian detection. Chen *et al.* [199] also get the best results with the "frequency encoding" among other encoding schemes for driver monitoring applications. Perot *et al.* [200] test different accumulation and encoding strategies for object detection, with the best results using the "discretized event volume" representation from Zhu *et al.* [201]. As a step above, [197] propose a "temporal active focus" representation, allowing them to surpass the results from other representations. Besides human-designed features, generalized expression can also be learned automatically in an *end-to-end* manner. Tulyakov *et al.* [202] 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 hand-

crafted approach. Cannici *et al.* [203] 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.* [204] apply a SNN to encode events and generate attention maps for further fusion with frame images, in an object detection framework. Or, as events are sparse, GNN (graph neural networks) can efficiently process events as spatio-temporal graphs with no information loss [205].

D. Applications in Autonomous Vehicles

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 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 the early stages, many published datasets (e.g. MVSEC [206], DDD17 [207], 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 [169]. Larger resolution benefits further object detection range and better recognition for small objects, while posing challenges for computation capability because of huge event flows. We expect to see more and more HD event camera datasets as Perot *et al.* [200] appear in public. The published datasets are for various purposes, such as target detection [207], lane detection [208], drowsiness detection [199] etc. A comprehensive summary of the current open datasets is demonstrated in Table III.

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 the DDD17 dataset, Chen *et al.* [198] 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.* [204], is used by Cao *et al.* [209] to detect vehicles. Hu *et al.* [210], [211] augment respectively parts of a day and a night sequence of MVSEC dataset [206] 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.* [204] 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.* [209] fuse events and frames at different encoding levels from parallel heads using feature attention gate components. Hu *et al.* [210], [211] 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 [212]–[215], in which individual datasets are utilized according to specific cameras. Chen *et al.* [212] compare different accumulation methods coupled with early fusion and late fusion schemes. In Jiang *et al.* [213], events and frames data channels are both fed into different CNNs and then fuse multiple confidence maps to achieve good pedestrian detection. Cladera *et al.* [214] implement a BNN (binary neural networks) on FPGA for fast detection. Wan *et al.* [215] propose a Pedestrian-SARI dataset and alternative events representations for asynchronous CNN detection. The lane extraction problem is investigated in Cheng *et al.* [208], 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.* [200], 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. GNN architecture of Schaefer *et al.* [205] is evaluated on this dataset and shows state-of-the-art results, while being low computationally intensive.

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 [216]–[219]. 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.* [216] group events into clusters via morphological operators, then track the multiple moving objects. Stoffregen *et al.* [217] warp the events several times to cluster moving objects. Zhou *et al.* [218] cluster the objects via graph cut on linked space-time event graph. [219] cluster the objects with split and merge strategy and track grouped events. Monda *et al.* [220] 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.* [221] focuses on drowsiness detection and compares some classification algorithms on their event dataset. Provided with a new dataset, Chen *et al.* [199] compares different CNN architectures and events accumulation schemes for driver drowsiness detection, gaze-zone, and hand-gesture recognition.

Finally, for all applications, a new trend will be to apply transformers networks on event data also, with both temporal and spatial attentions. It can even be used conjointly with SNN, such as Zhang *et al.* [222] in the case of object tracking.

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

TABLE I
SUMMARY OF VISION SENSOR FEATURES.

Visual Sensor	Characteristic								Applied Scenarios
	Resolution	Frame Rate	Night Vision	Adverse Weather	Specular Reflection	Dynamic Range	Cost	Technology Maturity	
RGB Camera	* * *	**	*	*	*	**	*	* * *	Object detection in good conditions General purpose
Near Infrared Camera*	* * *	**	**	**	**	**	*	* * *	Driver monitoring system Object detection at night
Thermal(LWIR) Camera	*	*	* * *	* * *	* * *	—	**	* * *	Object detection at night and adverse weather, living things detection
Range Gated Camera*	**	**	* * *	* * *	* * *	**	* * *	*	Object detection at night and adverse weather
Polarization Camera	* * *	**	*	**	* * *	* * *	**	*	Mitigate specular reflection, HDR imaging, object detection in adverse weather
Event Camera	**	* * *	**	*	*	* * *	**	**	Super fast object detection driver monitoring system

* Active sensor, used with a joint illuminator.

TABLE II
SEVERAL TYPICAL OPEN DATASETS FOR MULTIPLE SENSING MODALITIES (INFRARED, GATED INFRARED AND POLARIZATION CAMERAS).

Name	Modality						Size	Annotation	Location	Year
	RGB	NIR	LWIR	Gated	Polar	Event				
FLIR-ADAS [72]	✓		✓				13K	Person, car, bicycle dog, other vehicles	US	2020
C3I-ADAS [109]			✓				39K	Person, car, pole, bike, bicycle, bus	EU	2022
KAIST [104]	✓		✓				95K	Person, pedestrian and cyclist	Korea	2015
LLVIP [223]	✓		✓				18K	Pedestrian	China	2021
SCUT [105]			✓				211K	Walk/ride/squat people.	China	2019
ZUT [106]			✓				110K	9 classes including pedestrian, cyclist, animal	EU	2020
RANUS [224]	✓	✓					4K	10 classes including vehicle, road, pedestrian, vegetation	Korea	2018
SparsePPG [225]	✓	✓					19 seq.	With ground truth driver PPG waveform	US	2018
DENSE [226]	✓		✓	✓			13K	4 classes including pedestrian and car	EU	2020
PolarLITIS [162]	✓				✓		2.5K	Car, Person, bike, motorbike	EU	2021
ZJU-RGB-P [164]	✓				✓		394	Pixel-wise semantic segmentation building, glass, car, pedestrian, road, etc	China	2021
ViVID++ [227]	✓		✓			✓	22 driving seq.	Positions	Korea	2022

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.* [240]. 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 have the same dynamics, the event camera barely perceives the front car. Fortunately, such a problem could be overcome by applying RNN in object detection Perot *et al.* [200]. Other issues are the disturbances of shadows and of streetlights. The first is shown in Fig. 12 (d), where the shadow of a traffic sign on the ground may generate a false alarm. The second occurs when the scene is mostly illuminated by artificial lights (at night): most of them present a fast flicker effect invisible to the naked eye, but well visible to the event camera over enlightened scene. 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 deal with this problem, Khan *et al.* [241] propose an efficient compression algorithm.

VII. CONCLUSION AND FUTURE WORKS

Although RGB cameras have cost advantages and are extensively applied in current vehicles, their inherent limitations impede the deployment of autonomous driving systems beyond constrained ODDs. To overcome these drawbacks, other types of sensing modalities are emerging. In this paper, several of these sensors have been reviewed as complements to conventional RGB cameras, these include, infrared, range-gated, polarization, and event cameras. Among them, NIR cameras are being mass-produced and integrated into production cars. Some are still in the early stages, such as the polarization, event, and range-gated cameras. A concise summary of those sensors' characteristics and typical scenarios is given in Tab. I.

TABLE III
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 [194] *(includes previous PRED16 [193])*	240 × 180	Grey frames*	Mobile target following	1.25h	prey size, *prey position*	Northern Ireland	2018, *2016*
DDD20 [228] *(includes previous DDD17 [207])*	346 × 260	Grey frames* IMU* Car data GNSS	Vehicle control	39h *+ 12h*	-	USA, *Swiss, Germany*	2020, *2017*
PKU-DDD17-CAR† [204]			Detection		"Car"		2019
Ev-Seg† [229]			Segmentation	20 intervals	Semantic seg		2019
N-Cars [230]	304 × 240	-	Classification	24K samples	"Car" "Background"	unknown	2018
MVSEC [206]	346 × 260 2 cameras	Grey frames* Grey stereo camera IMU LiDAR GPS Motion capture	Depth Localisation	1h	Depth	USA	2018
MVSEC-OF‡ [231]			Optical flow		Optical flow		2018
MVSEC-DAY20‡ [210]			Detection	partial seq. "*outdoor_day2*"	"Car"		2020
MVSEC-NIGHTL21‡ [211]			Detection	partial seq. "*outdoor_night1*"	"Car"		2021
Slasher dataset [232]	346 × 260	Grey frames* Steering Radio localisation	Vehicle control	2 sequences	-	Swiss	2019
Event Camera Driving Sequences [233]	640 × 480	RGB camera	Frames reconstruction	40 sequences	-	Swiss	2019
CED [170]	346 × 260 RGB event cam	RGB frames*	Color frames reconstruction	50min	-	unknown	2019
Pedestrian Detection Dataset [234]	346 × 260 RGB event cam	-	Detection	12 recordings	"Pedestrian"	China	2019
EDDD§ [221]	346 × 260	-	Driver monitoring	260 sequences	Drowsiness	China	2020
NeuroIV§ [199]	346 × 260 RGB event cam	RGB frames* Depth maps NIR frames	Driver monitoring	27K samples	Drowsiness Gaze-zones Hand-gestures	China	2020
GAD Dataset [235]	304 × 240	-	Detection	39h	"Car" "Pedestrian"	France	2020
Brisbane Event VPR [236]	346 × 260 RGB event cam	RGB frames* RGB camera IMU* GPS	Visual place recognition	8km	Landmarks	Australia	2020
DENSE¶,‖ [237]	346 × 260	RGB frames Depth maps	Depth Segmentation	8K samples	Depth Semantic seg	-	2020
DSEC [159]	640 × 480 2 cameras	2x RGB cameras LiDAR§ RTK GPS§	Depth Localisation	53min	Depth	Swiss	2021
DSEC-OF** [238]			Optical flow		Optical flow		2021
EventScape¶ [239]	512 × 256	RGB frames Depth maps Car data	Depth Segmentation	2h	Depth Semantic seg	-	2021
Pedestrian-SARI§ [215]	346 × 260	Grey frames*	Detection	141 sequences	"Person"	China	2021
ViVID++ [227] *(driving scenes part)*	640 × 480	RGB camera Thermal camera LiDAR (for 8 seq.) RTK GPS	Localisation VSLAM	22 sequences	Positions	South Korea	2022
DET [208]	1280 × 800	-	Lane extraction	5h	Road lanes	China	2019
1Mp Detection [200]	1280 × 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 [226] presented in Table II and in Section IV-C.
** Extension of DSEC, providing ground truth to other problem.

These additional sensing modalities should enable the extension of the operating conditions of autonomous vehicles. The review has shown that most of the perception algorithms used for these sensors are similar to those used to process RGB images. That is, the RGB channels are replaced by infrared or polarization channels. For range-gated and event cameras, since their imaging principles are different, algorithms have been specifically designed successfully to leverage their unique imaging properties. Experiments in public roads have shown their advantages as well as their weaknesses. The need to ensure reliable and resilient perception systems for safety-critical vehicle maneuvers has led to the use of these sensors in conjunction with conventional sensors. The fusion between these emergent sensors and RGB cameras has resulted in several common fusion strategies. Field trials have demonstrated that these novel sensors can be part of different perception systems contributing to their performance. It is envisaged that lower prices, high reliability, and more powerful algorithms will be attained in the future, with some of these sensors playing critical roles in future ADAS or AD systems or for specific applications involving autonomous navigation in land, air, or sea.

ACKNOWLEDGMENT

This work has been carried out within SIVALab, joint laboratory between Renault and Heudiasyc UMR UTC/CNRS.

REFERENCES

[1] C. J. Cleveland and C. Morris, *Handbook of Energy*. Elsevier, 2013.
[2] A. Carrasco-Casado and R. Mata-Calvo, *Space Optical Links for Communication Networks*. Springer International Publishing, 2020, pp. 1057–1103.
[3] C. Urmson *et al.*, "Autonomous driving in urban environments: Boss and the Urban Challenge," *Journal of Field Robotics*, vol. 25, pp. 425–466, 2008.
[4] "Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles," SAE International, Tech. Rep., 2018.
[5] (2017) SIVALab. [Online]. Available: https://www.hds.utc.fr/partenariats/laboratoire-commun-sivalab.html
[6] 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.
[7] 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.
[8] C. Eising, J. Horgan, and S. Yogamani, "Near-field perception for low-speed vehicle automation using surround-view fisheye cameras," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, pp. 13 976 – 13 993, 2021.
[9] M. Gyllenhammar *et al.*, "Towards an operational design domain that supports the safety argumentation of an automated driving system," in *10th European Congress on Embedded Real Time Software and Systems*, 2020.
[10] E. Yurtsever, J. Lambert, A. Carballo, and K. Takeda, "A survey of autonomous driving: Common practices and emerging technologies," *IEEE Access*, vol. 8, pp. 58 443–58 469, 2020.
[11] J. Van Brummelen, M. OBrien, 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.
[12] G. Velasco-Hernandez, D. J. Yeong, J. Barry, and J. Walsh, "Autonomous driving architectures, perception and data fusion: A review," in *IEEE 16th International Conference on Intelligent Computer Communication and Processing (ICCP)*, 2020.
[13] Z. Wang *et al.*, "Multi-sensor fusion in automated driving: A survey," *IEEE Access*, vol. 8, pp. 2847–2868, 2019.
[14] E. Marti, M. A. de Miguel, F. Garcia, and J. Perez, "A review of sensor technologies for perception in automated driving," *IEEE Intelligent Transportation Systems Magazine*, vol. 11, pp. 94–108, 2019.
[15] K. M. Judd, M. P. Thornton, and A. A. Richards, "Review of spectral and polarization imaging systems," in *Proc. SPIE 11351, Unconventional Optical Imaging II*, 2020.
[16] C. Hughes, M. Glavin, E. Jones, and P. Denny, "Wide-angle camera technology for automotive applications: a review," *IET Intelligent Transport Systems*, vol. 3, p. 19 31, 2008.
[17] G. Gallego *et al.*, "Event-based vision: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, pp. 154–180, 2020.
[18] M. Heimberger *et al.*, "Computer vision in automated parking systems: Design, implementation and challenges," *Image and Vision Computing*, vol. 68, pp. 88–101, 2017.
[19] "Standard tables for reference solar spectral irradiances: Direct normal and hemispherical on 37 tilted surface," American Society for Testing Materials, Standard, 2012.
[20] J. D. Boullough *et al.*, "An investigation of headlamp glare: Intensity, spectrum and size," 2003.
[21] UNECE, "Regulation of no 112 of the economic commission for europe of the united nations (un/ece)," Tech. Rep., 2014.
[22] 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.
[23] 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.
[24] K. Weikl, D. Schroeder, and W. Stechele, "Optimization of automotive color filter arrays for traffic light color separation," in *Color and Imaging Conference*, 2020.
[25] C. Park *et al.*, "G-channel restoration for rwb cfa with double-exposed w channel," *Sensors*, vol. 17, pp. 8570–8594, 2017.
[26] OmniVision. (2019) Rgb-ir technology. [Online]. Available: https://www.ovt.com/purecel-pixel-tech/rgb-ir-technology/faqs
[27] M. Brown and S. Susstrunk, "Multispectral sift for scene category recognition," in *IEEE/CVF International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.
[28] M. Roser and P. Lenz, "Camera-based bidirectional reflectance measurement for road surface reflectivity classification," in *IEEE Intelligent Vehicles Symposium*, 2010.
[29] 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.
[30] S. Maddalena, D. A., and R. Diels, *Automotive CMOS Image Sensors*. springer, 2005.
[31] M. Bigasa, E. Cabrujaa, J. Forestb, and J. Salvi, "Review of cmos image sensors," *Microelectronics Journal*, vol. 37, pp. 433–451, 2006.
[32] N. Akahane and S. Sugawa, "Wide dynamic range cmos image sensors for high quality digital camera, security, automotive and medical applications," in *5th IEEE Conference on Sensors*, 2006.
[33] S. Gilroy, J. O'Dwyer, and L. Bortoleto, "Characterisation of cmos image sensor performance in low light automotive applications," *arXiv:2011.12436 [eess.IV]*, 2020.
[34] N. Stevanovic *et al.*, *Low-Cost High Speed CMOS Camera for Automotive Applications*. Springer, 2000.
[35] N. Behmann and H. Blume, "Real-time led flicker detection and mitigation: Architecture and fpga-implementation," in *25th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, 2018.
[36] M. Innocent *et al.*, "Automotive 8.3 mp cmos image sensor with 150 db dynamic range and light flicker mitigation," in *IEEE International Electron Devices Meeting (IEDM)*, 2021.
[37] (2017) SONY IMX390. [Online]. Available: https://www.sony-semicon.com/files/62/pdf/p-15_IMX390.pdf
[38] Y. Huang and Y. Chen, "Survey of state-of-art autonomous driving technologies with deep learning," in *IEEE 20th International Conference on Software Quality, Reliability and Security Companion*, 2020.
[39] K. He *et al.*, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *IEEE International Conference on Computer Vision*, 2015.
[40] J. Redmon *et al.*, "You only look once: Unified, real-time object detection," in *IEEE CVPR*, 2016.
[41] C.-Y. Wang *et al.*, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," *arXiv:2207.02696*, 2022.

[42] S. Ren *et al.*, "Faster r-cnn: towards real-time object detection with region proposal networks," in *Proceedings of the 28th International Conference on Neural Information Processing Systems*, 2015.

[43] W. Liu *et al.*, "SSD: Single Shot MultiBox Detector," in *Proceeding of European Conference on Computer Vision*, 2016.

[44] P. D. Varcheie and L. Gagnon, "Attention is all you need," in *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, 2017.

[45] N. Carion *et al.*, "End-to-end object detection with transformers," in *European Conference on Computer Vision*, 2020.

[46] R. Liu *et al.*, "End-to-end lane shape prediction with transformers," in *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2021.

[47] A. Prakash *et al.*, "Multi-modal fusion transformer for end-to-end autonomous driving," in *IEEE CVPR*, 2021.

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

[49] C. Fredembach and S. Susstrunk, "Colouring the near-infrared," in *Proceeding of 16th Color and Imaging Conference*, 2008.

[50] 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.

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

[52] E. de Borniol *et al.*, "High-performance 640 x 512 pixel hybrid ingaas image sensor for night vision," in *Proc. SPIE 8353, Infrared Technology and Applications XXXVIII*, 2012.

[53] Z. Chen, X. Wang, and R. Liang, "Rgb-nir multispectral camera," *Optics Express*, vol. 22, 2014.

[54] 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.

[55] C. Park and M. G. Kang, "Color restoration of rgbn multispectral filter array sensor images based on spectral decomposition," *Sensors*, vol. 16, no. 719, 2016.

[56] O. Skorka, P. Kane, and R. Ispasoiu, "Color correction for rgb sensors with dual-band filters for incabin imaging applications," in *Electronic Imaging, Autonomous Vehicles and Machines Conference*, 2019.

[57] 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.

[58] M. Dummer, K. Johnson, S. Rothwell, K. Tatah, and M. Hibbs-Brenner, "The role of vcsels in 3d sensing and lidar," in *Proc. SPIE 11692, Optical Interconnects XXI*, 2021.

[59] 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.

[60] 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.

[61] M. P. Hansen and D. S. Malchow, "Overview of SWIR detectors, cameras, and applications," in *Proc. SPIE 6939, Thermosense XXX*, 2008.

[62] X. Dai *et al.*, "Tirnet: Object detection in thermal infrared images for autonomous driving," *Applied Intelligence*, vol. 51, p. 12441261, 2020.

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

[64] R. Bhan *et al.*, "Uncooled infrared microbolometer arrays and their characterisation techniques," *Defence Science Journal*, vol. 59, pp. 580–589, 2009.

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

[66] J. Tissot *et al.*, "Uncooled microbolometer detector: recent developments at ulis," *Opto-Electronics Review*, vol. 14, pp. 25–32, 2006.

[67] J.-J. Yon *et al.*, "Latest amorphous silicon microbolometer developments at leti-lir," in *Proc. SPIE 6940, Infrared Technology and Applications XXXIV*, 2008.

[68] F. Niklaus *et al.*, "Mems-based uncooled infrared bolometer arrays: a review," in *Proc. SPIE 6836, MEMS/MOEMS Technologies and Applications III*, 2007.

[69] T. Oranowski, "Nonuniformity correction algorithm with efficient pixel offset estimation for infrared focal plane arrays," *SpringerPlus*, vol. 5, pp. 1831–1839, 2016.

[70] A. Tempellhahn, H. Budzier, and G. Gerlach, "Shutter-less calibration of uncooled infrared cameras," *Journal of Sensors and Sensor Systems*, vol. 5, pp. 9–16, 2016.

[71] C. Liu *et al.*, "Shutterless non-uniformity correction for the long-term stability of an uncooled long-wave infrared camera," *Measurement Science and Technology*, vol. 29, pp. 18–28, 2018.

[72] FLIR. (2019) FLIR thermal sensing for ADAS. [Online]. Available: https://www.flir.com/oem/adas/adas-dataset-form/

[73] J. Velzquez *et al.*, "Analysis of thermal imaging performance under extreme foggy conditions: Applications to autonomous driving," *journal of Imaging*, vol. 8, 2022.

[74] M. A. Farooq *et al.*, "Evaluation of thermal imaging on embedded gpu platforms for application in vehicular assistance systems," *arXiv:2201.01661*, 2022.

[75] N. Pinchon *et al.*, "All-weather vision for automotive safety: Which spectral band?" in *Advanced Microsystems for Automotive Applications*, 2018.

[76] 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. 16, pp. 2014–2027, 2015.

[77] L. Fridman, J. Lee, B. Reimer, and T. Victor, "Owl and lizard patterns of head pose and eye pose in driver gaze classification," *IET Computer Vision*, vol. 10, pp. 308–314, 2016.

[78] H. Yoon *et al.*, "Driver gaze detection based on deep residual networks using the combined single image of dual near-infrared cameras," *IEEE Access*, pp. 93 448 – 93 461, 2019.

[79] N. S. Martinelli and S. A. Boulanger, "Cadillac deville thermal imaging night vision system," in *SAE Technical Paper Series #2000-01-0323*, 2000.

[80] T. Tsuji, H. Hattori, M. Watanabe, and N. Nagaoka, "Development of night-vision system," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 3, 2002, pp. 203–209.

[81] Kyocera. (2022) Kyocera develops world's first automotive night vision system with white and near-infrared light diodes integrated into a single gan laser device. [Online]. Available: https://global.kyocera.com/newsroom/news/2022/000643.html

[82] J.-E. Kallhammer, "Night vision: requirements and possible roadmap for fir and nir systems," in *Proc. SPIE 6198, Photonics in the Automobile II*, 2006.

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

[84] 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.

[85] K. M. Judd, M. P. Thornton, and A. A. Richards, "Automotive sensing: Assessing the impact of fog on lwir, mwir, swir, visible and lidar imaging performance," in *Proc. SPIE 11002, Infrared Technology and Applications XLV*, 2019.

[86] A. Eskandarian, R. Sayed, P. Delaigue, J. Blum, and A. Mortazavi, "Advanced driver fatigue research," 2007.

[87] C. Ahlstrom, K. Kircher, and A. Kircher, "A gaze-based driver distraction warning system and its effect on visual behavior," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, pp. 965 – 973, 2013.

[88] C. Schwarz, J. Gaspar, T. Miller, and R. Yousefian, "The detection of drowsiness using a driver monitoring system," *Traffic Injury Prevention*, vol. 20, pp. 157–161, 2019.

[89] S. H. Park, H. S. Yoon, and K. R. Park, "Faster r-cnn and geometric transformation-based detection of drivers eyes using multiple near-infrared camera sensors," *Sensors*, vol. 19, pp. 1–29, 2019.

[90] D. Dinges *et al.*, "Evaluation of techniques for ocular measurement as an index of fatigue and the basis for alertness management," 1998.

[91] Q. Ji and X. Yang, "Real-time eye, gaze, and face pose tracking for monitoring driver vigilance," *Real-Time Imaging*, vol. 8, pp. 357–377, 2002.

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

[93] I. Garcia, S. Bronte, L. M. Bergasa, J. Almazan, and J. Yebes, "Vision-based drowsiness detector for real driving conditions," in *IEEE Intelligent Vehicles Symposium*, 2012.

[94] 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.

[95] R. A. Naqv, M. Arsalan, G. Batchuluun, H. S. Yoon, and K. R. Park, "Deep learning-based gaze detection system for automobile drivers using a nir camera sensor," *Sensors*, vol. 18, 2018.

[96] Y. Dong, Z. Hu, K. Uchimura, and N. Murayama, "Driver inattention monitoring system for intelligent vehicles: A review," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, pp. 596–614, 2010.

[97] A. A. Akinyelu and P. Blignaut, "Convolutional neural network-based methods for eye gaze estimation: A survey," *IEEE Access*, vol. 8, pp. 142 581–142 605, 2020.

[98] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *in Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2001.

[99] C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mininf and Knowledge Discovery*, vol. 2, pp. 121–167, June 1998.

[100] Y. Fang *et al.*, "A shape-independent method for pedestrian detection with far-infrared images," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 6, pp. 1679–1697, 2004.

[101] D. Forslund and J. Bjarkerfur, "Night vision animal detection," in *IEEE Intelligent Vehicles Symposium*, 2014.

[102] M. Kristo *et al.*, "Thermal object detection in difficult weather conditions using yolo," *IEEE Access*, pp. 125 459–125 476, 2020.

[103] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," *arXiv:1804.02767*, 2018.

[104] 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.

[105] Z. Xu *et al.*, "Benchmarking a large-scale fir dataset for on-road pedestrian detection," *Infrared Physics & Technology*, vol. 96, pp. 199–208, 2019.

[106] P. Tumas *et al.*, "Pedestrian detection in severe weather conditions," *IEEE Access*, vol. 8, pp. 62 775–62 784, 2020.

[107] M. A. Farooq *et al.*, "Object detection in thermal spectrum for advanced driver-assistance systems (adas)," *IEEE Access*, vol. 9, pp. 156 465 – 156 481, 2021.

[108] G. Jocher. (2021) Yolov5. [Online]. Available: https://github.com/ultralytics/yolov5

[109] M. A. Farooq *et al.* (2022) C3i thermal automotive dataset. [Online]. Available: https://ieee-dataport.org/documents/c3i-thermal-automotive-dataset

[110] J. Ma, Y. Ma, and C. Li, "Infrared and visible image fusion methods and applications: A survey," *Information Fusion*, vol. 45, pp. 153–178, 2019.

[111] H. Choi *et al.*, "Multi-spectral pedestrian detection based on accumulated object proposal with fully convolutional networks," in *23rd International Conference on Pattern Recognition (ICPR)*, 2016.

[112] 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.

[113] G. Humblot-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)*, 2020.

[114] K. Takumi *et al.*, "Multispectral object detection for autonomous vehicles," in *Proceedings of the on Thematic Workshops of ACM Multimedia*, 2017.

[115] 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.

[116] J. Liu *et al.*, "Multispectral deep neural networks for pedestrian detection," in *Proceedings of the British Machine Vision Conference (BMVC)*, 2016.

[117] C. Li *et al.*, "Illumination-aware faster r-cnn for robust multispectral pedestrian detection," *Pattern Recognition*, vol. 85, pp. 161–171, 2019.

[118] ——, "Multispectral pedestrian detection via simultaneous detection and segmentation," in *In Proceeding of British Machine Vision Conference*, 2018.

[119] 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.

[120] 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.

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

[122] K. Dasgupta *et al.*, "Spatio-contextual deep network based multimodal pedestrian detection for autonomous driving," *arXiv:2105.12713*, 2021.

[123] Q. Fang, D. Han, and Z. Wang, "Cross-modality fusion transformer for multispectral object detection," in *arXiv:2111.00273*, 2022.

[124] Y. Zhang *et al.*, "Attention based multi-layer fusion of multispectral images for pedestrian detection," *IEEE Access*, vol. 8, pp. 2169–3536, 2020.

[125] Q. Jiang *et al.*, "Attention-based cross-modality feature complementation for multispectral pedestrian detection," *IEEE Access*, vol. 10, pp. 2169–3536, 2022.

[126] K. Dasgupta *et al.*, "Spatio-contextual deep network-based multimodal pedestrian detection for autonomous driving," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, pp. 15 940–15 950, 2022.

[127] L. Gillespie, "Apparent illuminance as a function of range in gated, laser night-viewing systems," *Journal of the Optical Society of America*, vol. 56, pp. 883–887, 1966.

[128] I. M. Baker *et al.*, "A low-noise laser-gated imaging system for long-range target identification," in *Proc. SPIE 5406, Infrared Technology and Applications*, 2004.

[129] A. M. Pinto and A. C. Matos, "Maresye: A hybrid imaging system for underwater robotic applications," *Information Fusion*, vol. 55, pp. 16–29, 2020.

[130] 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.

[131] O. David, N. Kopeika, and B. Weizer, "Range gated active night vision system for automobiles," *Applied Optics*, 2006.

[132] N. Spooren *et al.*, "RGB-NIR active gated imaging," in *Electro-Optical and Infrared Systems: Technology and Applications XIII*, 2016.

[133] A. H. Willitsford *et al.*, "Range-gated active short-wave infrared imaging for rain penetration," *Optical Engineering*, vol. 60, no. 1, pp. 1 – 11, 2021.

[134] F. Rutz *et al.*, "Ingaas apd matrix sensors for swir gated viewing," *Advanced Optical Technologies*, vol. 8, pp. 445–450, 2019.

[135] S. Burri *et al.*, "Architecture and applications of a high resolution gated spad image sensor," *Optics Express*, vol. 22, pp. 17 573–17 589, 2014.

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

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

[138] S. Walz, T. Gruber, W. Ritter, and K. Dietmayer, "Uncertainty depth estimation with gated images for 3d reconstruction," in *IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, 2020.

[139] F. Christnacher, J.-M. Poyet, M. Laurenzis, J.-P. Moegline, and F. Taillade, "Bistatic range-gated active imaging in vehicles with leds or headlights illumination," in *Proc. SPIE 7675, Photonics in the Transportation Industry: Auto to Aerospace III*, 2010.

[140] Y. Grauer, "Active gated imaging in driver assistance system," *Advanced Optical Technologies*, vol. 3, pp. 151–160, 2014.

[141] Y. Grauer and E. Sonn, "Active gated imaging for automotive safety applications," in *Proc. SPIE Video Surveillance and Transportation Imaging Applications*, 2015.

[142] F. Julca-Aguilar, J. Taylor, M. Bijelic, F. Mannan, E. Tseng, and F. Heide, "Gated3d: Monocular 3d object detection from temporal illumination cues," *arXiv:2102.03602*, 2021.

[143] G. Tobias, F. Julca-Aguilar, M. Bijelic, and F. Heide, "Gated2depth: Real-time dense lidar from gated images," in *The IEEE International Conference on Computer Vision (ICCV)*, 2019.

[144] Lucid Vision Labs. (2018) Beyond conventional imaging: Sony's polarized sensor. [Online]. Available: https://thinklucid.com/tech-briefs/polarization-explained-sony-polarized-sensor/

[145] Polarizer: Wikipedia. (2021) Polarizer. [Online]. Available: https://en.wikipedia.org/wiki/Polarizer

[146] J. J. Foster *et al.*, "Polarisation vision: overcoming challenges of working with a property of light we barely see," *The Science of Nature*, vol. 105, no. 27, pp. 1–26, 2018.

[147] D. Kliger, *Polarized Light in Optics and Spectroscopy*. Academic Press, 1990.

[148] M. Bass, E. V. Stryland, D. Williams, and W. Wolfe, *Handbook of Optics*. McGraw-Hill, 1996.

[149] O. Morel, F. Meriaudeau, C. Stolz, and P. Gorria, "Polarization imaging applied to 3d reconstruction of specular metallic surfaces," in *Proc. SPIE 5679, Machine Vision Applications in Industrial Inspection XIII*, 2005, pp. 178–186.

[150] L. B. Wolff, "Polarization camera for computer vision with a beam splitter," *Journal of the Optical Society of America A*, vol. 11, pp. 2935–2945, 1994.

[151] Y. Li *et al.*, "Multiframe-based high dynamic range monocular vision system for advanced driver assistance systems," *IEEE Sensors Journal*, vol. 15, pp. 5433–5441, 2015.

[152] Y. Wang *et al.*, "Efficient road specular reflection removal based on gradient properties," *Multimedia Tools and Applications*, vol. 77, pp. 30 615–30 631, 2018.

[153] E. Yamaguchi *et al.*, "Glass detection using polarization camera and lrf for slam in environment with glass," in *21st International Conference on Research and Education in Mechatronics (REM)*, 2022.

[154] F. Wang, S. Ainouz, F. Meriaudeau, and A. Bensrhair, "Polarization-based car detection," in *25th IEEE International Conference on Image Processing (ICIP)*, 2018.

[155] M. Sheeny, A. Wallace, M. Emambakhsh, S. Wang, and B. Connor, "POL-LWIR vehicle detection: Convolutional neural networks meet polarised infrared sensors," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2018.

[156] F. Wang, S. Ainouz, C. Petitjean, and A. Bensrhair, "Specularity removal: A global energy minimization approach based on polarization imaging," *Computer Vision and Image Understanding*, vol. 158, pp. 31–39, 2017.

[157] X. Wu *et al.*, "Hdr reconstruction based on the polarization camera," *IEEE Robotics and Automation Letters*, vol. 5, pp. 5113–5119, 2020.

[158] C. Posch, T. Serrano-Gotarredona, B. Linares-Barranco, and T. Delbruck, "Retinomorphic event-based vision sensors: Bioinspired cameras with spiking output," *Proceedings of the IEEE*, vol. 102, no. 10, pp. 1470–1484, 2014.

[159] M. Gehrig, W. Aarents, D. Gehrig, and D. Scaramuzza, "DSEC: A stereo event camera dataset for driving scenarios," *IEEE Robotic and Automation Letters*, 2021.

[160] R. Blin, S. Ainouz, S. Canu, and F. Meriaudeau, "Road scenes analysis in adverse weather conditions by polarization-encoded images and adapted deep learning," in *IEEE Intelligent Transportation Systems Conference (ITSC)*, 2019.

[161] ——, "A new multimodal rgb and polarimetric image dataset for road scenes analysis," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2020.

[162] ——, "The polarlitis dataset: Road scenes under fog," *IEEE Transactions on Intelligent Transportation Systems*, p. Early Access, 2021.

[163] M. Blanchon *et al.*, "Outdoor scenes pixel-wise semantic segmentation using polarimetry and fully convolutional network," in *14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 2019.

[164] K. Xiang, K. Yang, and K. Wang, "Polarization-driven semantic segmentation via efficient attention-bridged fusion," *Optics Express*, vol. 29, pp. 4802–4820, 2021.

[165] G. Gallego, T. Delbruck, G. M. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A. Davison, J. Conradt, K. Daniilidis, and D. Scaramuzza, "Event-based vision: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 154–180, 2020.

[166] 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.

[167] T. Finateu *et al.*, "5.10 a 1280720 back-illuminated stacked temporal contrast event-based vision sensor with 4.86m pixels, 1.066geps readout, programmable event-rate controller and compressive data-formatting pipeline," in *IEEE International Solid- State Circuits Conference*, 2020.

[168] Y. Suh *et al.*, "A 1280960 dynamic vision sensor with a 4.95-m pixel pitch and motion artifact minimization," in *IEEE International Symposium on Circuits and Systems*, 2020.

[169] S. Chen and M. Guo, "Live demonstration: Celex-v: A 1m pixel multi-mode event-based sensor," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019.

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

[171] N. Qiao, H. Mostafa, F. Corradi, M. Osswald, F. Stefanini, D. Sumislawska, 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.

[172] F. Akopyan *et al.*, "TrueNorth: Design and tool flow of a 65 mW 1 million neuron programmable neurosynaptic chip," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 34, no. 10, pp. 1537–1557, 2015.

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

[174] 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.

[175] 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.

[176] 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.

[177] 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.

[178] T. Barbier, C. Teuliere, 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.

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

[180] F. Paredes-Valls, K. Y. W. Scheper, and G. C. H. E. d. Croon, "Unsupervised learning of a hierarchical spiking neural network for optical flow estimation: From events to global motion perception," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 8, pp. 2051–2064, 2020.

[181] C. M. Parameshwara, S. Li, C. Fermller, N. J. Sanket, M. S. Evanusa, and Y. Aloimonos, "SpikeMS: Deep spiking neural network for motion segmentation," *arXiv:2105.06562*, 2021.

[182] R. Kreiser, "A neuromorphic approach to path integration: A head-direction spiking neural network with vision-driven reset," in *IEEE International Symposium on Circuits and Systems*, 2018.

[183] R. Kreiser *et al.*, "Error estimation and correction in a spiking neural network for map formation in neuromorphic hardware," in *IEEE International Conference on Robotics and Automation*, 2020.

[184] R. Stagsted, A. Vitale, J. Binz, A. Renner, L. Bonde Larsen, and Y. Sandamirskaya, "Towards neuromorphic control: A spiking neural network based PID controller for UAV," in *Robotics: Science and Systems XVI*, 2020.

[185] A. Vitale, A. Renner, C. Nauer, D. Scaramuzza, and Y. Sandamirskaya, "Event-driven vision and control for UAVs on a neuromorphic chip," *arXiv:2108.03694*, 2021.

[186] B. Rckauer, N. Knzig, S.-C. Liu, T. Delbruck, and Y. Sandamirskaya, "Closing the accuracy gap in an event-based visual recognition task," *arXiv:1906.08859 [cs]*, 2019.

[187] N. Salvatore, S. Mian, C. Abidi, and A. D. George, "A neuro-inspired approach to intelligent collision avoidance and navigation," in *AIAA/IEEE 39th Digital Avionics Systems Conference*, 2020.

[188] V. Brebion, J. Moreau, and F. Davoine, "Real-time optical flow for vehicular perception with low- and high-resolution event cameras," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 9, pp. 15 066–15 078, 2022.

[189] D. Zou *et al.*, "Robust dense depth maps generations from sparse DVS stereos," in *British Machine Vision Conference*, 2017.

[190] H. Rebecq, T. Horstschaefer, and D. Scaramuzza, "Real-time visual-inertial odometry for event cameras using keyframe-based nonlinear optimization," in *British Machine Vision Conference*, 2017.

[191] D. Joubert, M. Hbert, H. Konik, and C. Lavergne, "Characterization setup for event-based imagers applied to modulated light signal detection," *Applied Optics*, vol. 58, no. 6, pp. 1305–1317, 2019.

[192] H. Akolkar *et al.*, "What can neuromorphic event-driven precise timing add to spike-based pattern recognition?" *Neural Computation*, vol. 27, no. 3, pp. 561–593, 2015.

[193] D. P. Moeys *et al.*, "Steering a predator robot using a mixed frame/event-driven convolutional neural network," in *Second Interna-*

tional Conference on Event-based Control, Communication, and Signal Processing, 2016.

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

[195] 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.

[196] F. Paredes-Valls, J. Hagenaars, and G. de Croon, "Self-supervised learning of event-based optical flow with spiking neural networks," *arXiv:2106.01862*, 2021.

[197] B. Liu, "Motion robust high-speed light-weighted object detection with event camera," 2022.

[198] 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.

[199] G. Chen *et al.*, "NeuroIV: Neuromorphic vision meets intelligent vehicle towards safe driving with a new database and baseline evaluations," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–13, 2020.

[200] E. Perot, P. de Tournemire, D. Nitti, J. Masci, and A. Sironi, "Learning to detect objects with a 1 megapixel event camera," *Advances in Neural Information Processing Systems*, vol. 33, pp. 16 639–16 652, 2020.

[201] A. Z. Zhu *et al.*, "Unsupervised event-based learning of optical flow, depth, and egomotion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019.

[202] S. Tulyakov, F. Fleuret, M. Kiefel, P. Gehler, and M. Hirsch, "Learning an event sequence embedding for dense event-based deep stereo," in *IEEE/CVF International Conference on Computer Vision*, 2019.

[203] M. Cannici, M. Ciccone, A. Romanoni, and M. Matteucci, "A differentiable recurrent surface for asynchronous event-based data," in *European Conference on Computer Vision*, 2020.

[204] J. Li, S. Dong, Z. Yu, Y. Tian, and T. Huang, "Event-based vision enhanced: A joint detection framework in autonomous driving," in *IEEE International Conference on Multimedia and Expo*, 2019.

[205] S. Schaefer, D. Gehrig, and D. Scaramuzza, "Aegnn: Asynchronous event-based graph neural networks," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 12 361–12 371.

[206] A. Z. Zhu, D. Thakur, T. zaslan, B. Pfrommer, V. Kumar, and K. Daniilidis, "The multivehicle stereo event camera dataset: An event camera dataset for 3d perception," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2032–2039, 2018.

[207] J. Binas, D. Niel, S.-C. Liu, and T. Delbruck, "DDD17: End-to-end DAVIS driving dataset," *Workshop on Machine Learning for Autonomous Vehicles*, 2017.

[208] W. Cheng *et al.*, "DET: A high-resolution DVS dataset for lane extraction," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019.

[209] H. Cao *et al.*, "Fusion-based feature attention gate component for vehicle detection based on event camera," *IEEE Sensors Journal*, vol. 21, no. 21, pp. 24 540–24 548, 2021.

[210] Y. Hu, T. Delbruck, and S.-C. Liu, "Learning to exploit multiple vision modalities by using grafted networks," in *European Conference on Computer Vision*, 2020.

[211] Y. Hu, S.-C. Liu, and T. Delbruck, "v2e: From video frames to realistic dvs events," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, June 2021.

[212] G. Chen *et al.*, "Multi-cue event information fusion for pedestrian detection with neuromorphic vision sensors," *Frontiers in Neurorobotics*, vol. 13, pp. 10–16, 2019.

[213] Z. Jiang *et al.*, "Mixed frame-/event-driven fast pedestrian detection," in *International Conference on Robotics and Automation*, 2019.

[214] F. C. Ojeda, A. Bisulco, D. Kepple, V. Isler, and D. D. Lee, "On-device event filtering with binary neural networks for pedestrian detection using neuromorphic vision sensors," in *IEEE International Conference on Image Processing*, 2020.

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

[216] A. Mitrokhin, C. Fermller, C. Parameshwara, and Y. Aloimonos, "Event-based moving object detection and tracking," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2018.

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

[218] 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.

[219] C. M. Parameshwara *et al.*, "0-MMS: Zero-shot multi-motion segmentation with a monocular event camera," in *IEEE International Conference on Robotics and Automation*, 2021.

[220] A. Mondal *et al.*, "Moving object detection for event-based vision using graph spectral clustering," in *IEEE/CVF International Conference on Computer Vision Workshops*, 2021.

[221] 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.

[222] J. Zhang, B. Dong, H. Zhang, J. Ding, F. Heide, B. Yin, and X. Yang, "Spiking transformers for event-based single object tracking," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 8791–8800.

[223] X. Jia *et al.*, "LLVIP: A visible-infrared paired dataset for low-light vision," in *IEEE/CVF International Conference on Computer Vision Workshops*, 2021.

[224] 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.

[225] E. M. Nowara *et al.*, "Sparseppg: Towards driver monitoring using camera-based vital signs estimation in near-infrared," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018.

[226] DENSE. (2020) Dense dataset. [Online]. Available: https://www.uni-ulm.de/en/in/driveu/projects/dense-datasets/

[227] A. J. Lee, Y. Cho, Y.-s. Shin, A. Kim, and H. Myung, "Vivid++: Vision for visibility dataset," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 6282–6289, 2022.

[228] 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.

[229] 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.

[230] 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.

[231] 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.

[232] 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.

[233] 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.

[234] S. Miao, G. Chen, X. Ning, Y. Zi, K. Ren, Z. Bing, and A. Knoll, "Neuromorphic vision datasets for pedestrian detection, action recognition, and fall detection," *Frontiers in Neurorobotics*, vol. 13, 2019.

[235] P. de Tournemire, "A large scale event-based detection dataset for automotive," *arXiv:2001.08499*, 2020.

[236] 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.

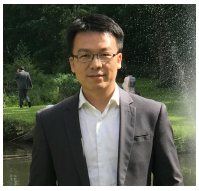
[237] D. G. Javier Hidalgo-Carrio and D. Scaramuzza, "Learning monocular dense depth from events," *IEEE International Conference on 3D Vision*, 2020.

[238] M. Gehrig, M. Millhäusler, D. Gehrig, and D. Scaramuzza, "E-raft: Dense optical flow from event cameras," in *International Conference on 3D Vision*, 2021.

[239] D. Gehrig, M. Regg, 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.

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

[241] 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.



You LI is a senior research engineer at the Groupe RENAULT. He received a B.Sc degree (2006) and M.Sc degree (2010) from Nanjing University of Science and Technology, and Shanghai Jiao Tong University, respectively. He completed his Ph.D thesis on machine perception applied to intelligent vehicles (2013) at the Universit de Technologie de Belfort Montbliard, France. From 2014, he worked as a post-doc researcher with the CNRS (French National Center for Scientific Research) on unmanned surface vehicle. In 2016, he joined TNO (Netherlands Organization for Applied Scientific Research) as a research scientist working on sensor fusion for autonomous vehicles. Since 2017, he is with research division of Groupe RENAULT. His main research interests are in LiDAR based perception systems for autonomous vehicles.

Julien Moreau received the Ph.D. degree in computer vision from the University of Technology of Belfort-Montbliard (UTBM), France, in 2016. He accomplished postdoctoral positions in The French Institute of Science and Technology for Transport Development and Networks (IFSTTAR), in Lille, France, and in the Institute of Information and Communication Technologies Electronics and Applied Mathematics (ICTEAM), at Universit Catholique de Louvain, Louvain-la-Neuve, Belgium. Since 2019, he is an associate professor in the Computer Science department of University of Technology of Compigne (UTC), France, and is carrying out his research in Heudiasyc UMR 7253, a joint UTC-CNRS research laboratory. From that, he is also a member of SIVALab, a joint laboratory between Renault, UTC and CNRS. His research interests cover stereovision, unconventional cameras, calibration and machine learning applied to perception and localization for mobile robotics.

Javier Ibanez-Guzman (Member, IEEE) received the M.S.E.E. degree from the University of Pennsylvania, USA, as a Fulbright Scholar, and the Ph.D. degree from the University of Reading on an UK SERC fellowship. He was Visiting Scholar at the University of California, Berkeley. Currently Corporate Expert on Autonomous Systems at Renault S.A., and co-director of the SIVALab Common Laboratory between the CNRS, UTC Compiegne and Renault working on intelligent vehicle technologies. Senior Editor and Associated Editor for related IEEE Transactions as well as representative to ISO groups associated to autonomous vehicles and AI. He is expert to the EU and Eureka research programmes. Formerly, he was a Senior Scientist with SimTech, A-Star research institute, Singapore, where he spearheaded work on autonomous ground vehicles. He is a C.Eng. and fellow of the Institute of Engineering Technology, U.K.