Article

Embodied neuromorphic synergy for lighting-robust machine vision to see in extreme bright

Shijie Lin [1], Guangze Zheng[1], Ziwei Wang[2], Ruihua Han[1], Wanli Xing[1], Zeqing Zhang [1], Yifan Peng [1,3] ✉ & Jia Pan [1,4] ✉

Proper exposure settings are crucial for modern machine vision cameras to accurately convert light into clear images. However, traditional auto-exposure solutions are vulnerable to illumination changes, splitting the continuous acquisition of unsaturated images, which significantly degrades the overall performance of underlying intelligent systems. Here we present the neuromorphic exposure control (NEC) system. This system effectively alleviates the longstanding saturation problem at its core by exploiting bio-principles found in peripheral vision to compute a trilinear event double integral (TEDI). This approach enables accurate connections between events and frames in the physics space for swift irradiance prediction, ultimately facilitating rapid control parameter updates. Our experimental results demonstrate the remarkable efficiency, low latency, superior generalization capability, and bio-inspired nature of the NEC in delivering timely and robust neuromorphic synergy for lighting-robust machine vision across a wide range of real-world applications. These applications encompass autonomous driving, mixed-reality, and three-dimensional reconstruction.

Modern machine vision cameras, or more generally referred to as active pixel sensors (APS), convert light into electrical signals through accumulated exposure, delivering continuous images and essentially serve as the primary visual input of various intelligent systems, such as autonomous driving[1], surgical robots[2], and mixed reality[3]. However, the rapid expansion of the market (projected to reach 32.2 billion USD by 2029[4]) is accompanied by vulnerabilities to illumination changes. In the absence of appropriate settings, sensor measurements may suffer from over/under-exposure, splitting the continuous acquisition of unsaturated images that significantly degrade the overall performance of underlying intelligent systems. For instance, in visual optometry[5,6], the presence of saturated images hinders the effective connections between frames, distorts trajectories, and corrupts the 3D modeling process. Saturated images also blind the sensor, resulting in the loss of effective perception, thereby introducing bias into the decision-

making process. An illustrative instance of this problem is the failures[7] of emergency braking in autonomous vehicles under a sudden transition from dark to bright. To mitigate these saturation problems, cameras require the capability to adjust their attributes for data acquisition, a task commonly referred to exposure control (EC).

The research on exposure controls (EC), as depicted in Fig. 1a, primarily relies on image sampling. Early approaches rely on simple image statistics like average intensity[8,9] or histogram[10,11] to solve the exposure control as a feedback control problem. However, their convergence to proper exposure consumes many image samples and requires an even distribution of scene illumination, making the adjustment slow in natural scenes. When strong lighting variations happen, these methods generally take a long time to help the camera adapt to the new irradiance level. Later works reformulate the problem as a prediction problem and exploit information including image

[1]Department of Computer Science, The University of Hong Kong, Pokfulam Rd, Hong Kong SAR, China. [2]College of Engineering and Computer Science Australian National University, Canberra, ACT, Australia. [3]Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam Rd, Hong Kong SAR, China. [4]LimX Dynamics, Shenzhen, China. ✉e-mail: evanpeng@hku.hk; jpan@cs.hku.hk
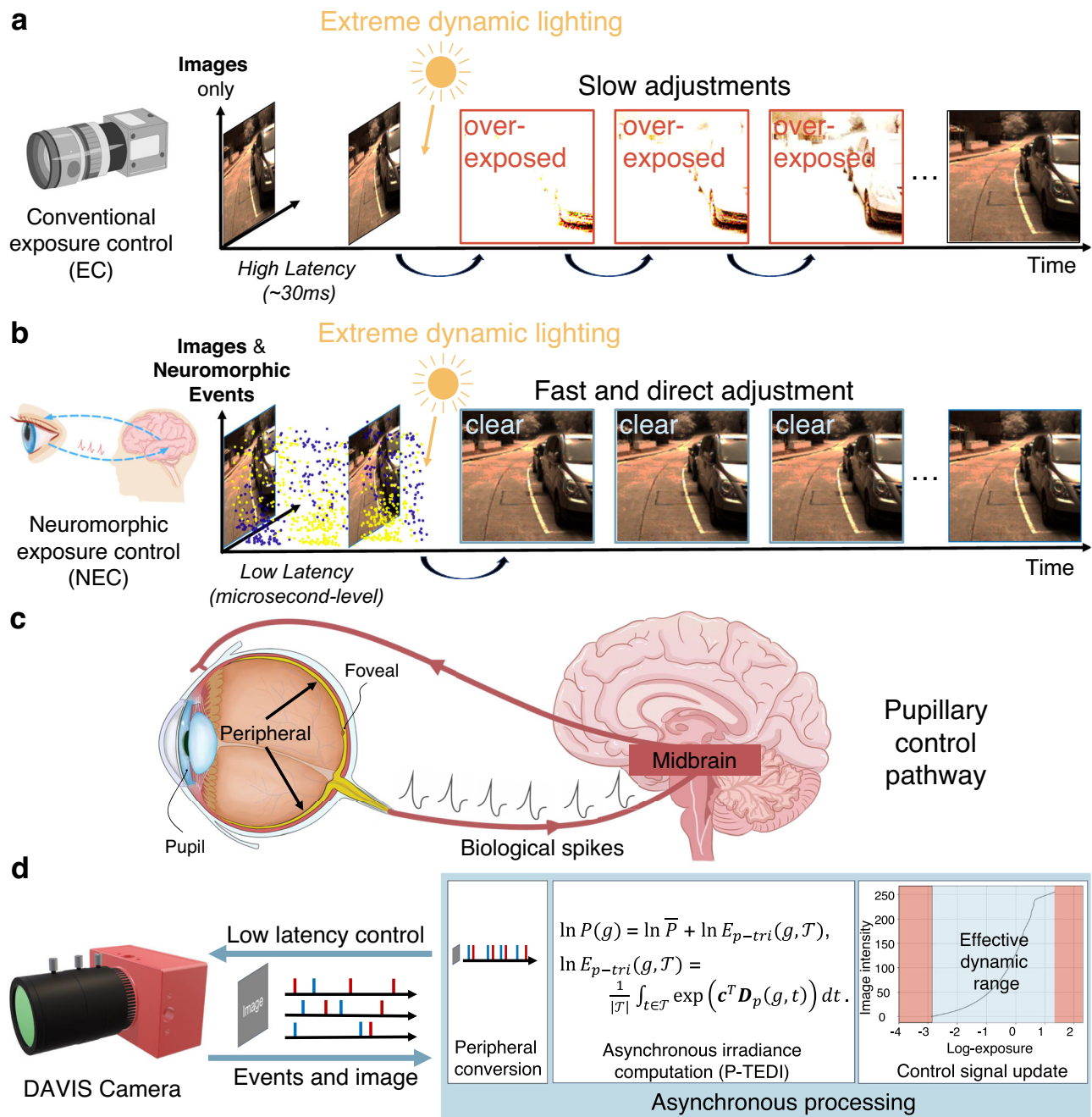
Fig. 1 | Principle and system workflow of the NEC. Diagrams of (**a**) conventional exposure control (EC) and (**b**) proposed neuromorphic exposure control (NEC) with an analogy of (**d**) neuromorphic exposure control to (**c**) pupil control pathway. The NEC brings fundamental bio-principles from the pupillary control pathway and peripheral vision into the design space and implementation, offering low latency, efficiency, and simple yet effective exposure control, which fundamentally avoids over/under-exposure (shown in **b**). In contrast, the conventional exposure control adjusts parameters slowly, resulting in over-exposed images under extreme dynamic lighting (shown in **a**).

entropy[12], image gradient[13], image features[6,14] semantics[15–17], and optical flow[18] in captured images to predict exposure setting for future capturing. However, the success prediction highly relies on the quality of the image data and the context in the scenes. For low-level statistics like gradient[13] or entropy[12], its performance is easily degraded by motion artifacts, which happen commonly in robotics systems when there are relative motions. For high-level semantics, like object consistency[17], one cannot guarantee the scenes always contain objects that satisfy the need of the control pipeline.

Currently, previous discussions are limited to the image data. Consequently, the updates of camera parameters only occur after the saturation has taken place. This is a strong circular dependency that fundamentally leads to the saturation problem. Furthermore, natural lighting is intricate and encompasses a dynamic range exceeding 280 bB[19,20] (requiring >1.2 Million digits for representation). This implies that significant changes in natural lighting cannot be captured through single sampling, particularly when employing modern machine vision cameras with a dynamic range of only 60 bB (equivalent to 256 digits for 8-bit images). As a result, standard built-in EC[21] typically consumes dozens of images to search for the optimal exposure parameter. The consequence is the presence of a series of saturated images, as commonly observed under lighting changes (e.g.,

Fig. 1a). Moreover, the framing process incurs a substantial latency (>30 ms for a 30 Hz camera), and when multiple image samples are employed, the overall latency accumulates linearly, significantly impairing performance of the intelligent system for a considerable duration (hundreds of milliseconds). Therefore, it is imperative to develop fundamental solutions that break the circular dependency and effectively handle rapid illumination changes encountered in real-world scenarios.

Human eyes possess remarkable capabilities to function effectively under challenging lighting and motions. These capabilities are governed by important bio-principles that hold great potential for inspiring the performance improvement of machine vision cameras in extreme lighting scenarios. In the human midbrain, a pupillary control pathway[22] (shown in Fig. 1c) has been established to leverage the low-latency biological spiking signals generated from the retina, where rod and cone cells are located, to regulate the total amount of light exposed to our eyes. Drawing inspiration from this bio-principle, we have incorporated it into the development of EC algorithms. As illustrated in Fig. 1b, our approach, dubbed as NEC, combines neuromorphic events and image modalities in an asynchronous manner to continuously update the exposure parameters. By simulating biological spikes from the rods and cones, the neuromorphic events exhibit high sensitivity to even minor illumination changes, operating independently of the exposure setting. This characteristic enables rapid updates of the control signal and fundamentally breaks the circular dependency inherent in traditional EC solutions.

The system design of NEC, as illustrated in Fig. 1d, takes both events (provided by the event camera[23,24]) and frames as input, connecting them in physical space to provide a high-rate update of the scene irradiance, which represents the scene illumination. To reversely compute an optimal exposure setting (see "Control framework"), we employ the camera response function[25] (CRF). However, such computation is highly inefficient for the nonlinear event generation model, requiring dense recursive computation event-wise in quadratic time complexity[26] (see "Physical relation for events and frames"). In contrast, the peripheral vision in human eyes[27,28] (see Fig. 1c, the eye) operates at a lower resolution to encompass a wider field of view, effectively reducing the computational load on the brain by nearly $350 \times$ [29] while still providing a general understanding of the scene. We utilize this bio-principle in human eyes to develop the peripheral conversion, which first converts events and frames to a resolution-reduced peripheral vision space and then computes the peripheral version of the proposed trilinear event double integral (P-TEDI, see "Fast irradiance computation in peripheral vision"). This approach significantly improves the overall computation speed (600 × improvements, Runtime evaluation) while maintaining accuracy (see proof in Supplementary Section 3), satisfying real-time constraints imposed by the EC tasks (real-time factor equal to 19, Runtime evaluation).

Our proposed NEC demonstrates embodied neuromorphic synergy[30,31] for modern machine vision cameras in various practical applications, including autonomous driving (Object detection for safe autonomous driving), mixed reality (In-field hand-pose detection for mixed reality), 3D reconstruction (Feature detection for finer 3D reconstruction), medical facial recognition (Sectio). Results indicate that with the proposed NEC, core algorithms (object detection[32], hand recognition[33], feature tracking[34]) in these systems consistently achieve higher accuracy and better performance under extreme lighting changes. In addition, the NEC exhibits a high level of generalizability, successfully avoiding saturation problems in various scenes (urban, highway, campus, and laboratory), motions (static, head moving, handheld shaking, and vehicles driving), and lighting conditions (indoor lighting, direct sunlight, shadows, reflections, and tunnel lighting). Notably, in the static indoor evaluation (see Supplementary Section 4.6), the number of saturated images is reduced from

hundreds (355 for Navid[10]) to only one slightly saturated image (around 10 percent pixels). The high efficiency, low latency, and bio-inspired nature of the proposed NEC provide timely and robust neuromorphic synergy for lighting-robust machine vision at a low cost. This can benefit a substantial number of real-world intelligent applications and guide the design of next-generation machine vision systems.

Results

We assess the proposed neuromorphic exposure control (NEC) in both simulated and real-world tasks, including object detection, hand/face key-point detection, feature point tracking, and 3D reconstruction across various scenes (urban tunnel, residential, highway, campus, hospital, and indoor laboratory), lighting conditions (tunnel/building/laboratory lighting and direct sunlight), and platforms (vehicle/head-mounted/handheld systems).

In the evaluation, we compare the NEC with state-of-the-arts, including Navid[10], Shim[13], and Begin[18]. Navid[10] is a widely used image histogram for feedback exposure control. Shim[13] is a prediction-based method that maximizes the gradient in the synthetic images to provide robust predictions. Begin[18] is based on the optical flow and saturation evaluation for better exposure control. We provide the semantics of three methods in the Supplementary Fig. S9 for reference. Three methods are implemented in C++ with optimized parameters by their authors to ensure efficiency. To ensure robustness, we disable the gain for Navid in real-world experiments. The maximum exposure time and gain settings for each experiment can be found in Supplementary Table S2. Additionally, in the evaluations of simulated driving in "Synthetic evaluation in CARLA simulator", we compare the NEC with fixed exposure, which involves setting the parameters at the beginning and keeping them unchanged throughout each test.

Object detection for safe autonomous driving

Ensuring safety is of utmost importance in autonomous driving[35,36]. However, conventional vision sensors often struggle to quickly adjust the exposure parameter when encountering scenes with significant illumination changes. This results in a *blind effect* phenomenon, where the vision system fails to detect critical objects and misses out important information for subsequent decision-making. Such challenges are frequently encountered in typical driving scenarios, such as entrance zones of tunnels or parking garages. The NEC overcomes these limitations by providing rapid response at the microsecond level. It effectively reduces saturation and ensures effective and reliable operation of the sensing system.

Synthetic evaluation in CARLA simulator. Experimental Setup. We simulate the entire process of image acquisition from a vehicle driving at a scene with significant illumination changes using the CARLA simulator[37]. As shown in Fig. 2a, the driving scenarios consist of urban cities during the daytime. We utilize the *Tesla Model 3* as the test vehicle and select the tunnel environment in *Town 03*. The route is traversed twice, starting from the middle of the tunnel and ending at two different sides of the tunnel, respectively. The vehicle exits the tunnel at an average speed of 200 km h^{-1}. The illumination change between the inner and outer of the tunnel is set to 120 bB, replicating a typical midday sunlight. Consequently, the on-vehicle camera will experience strong illumination changes as the vehicle drives out of the tunnel. Toward the end of the route, we place several cars and pedestrians to assess the detection performance under various exposure control algorithms. For more detailed information, please refer to Supplementary Section 4.2.

Object Detection. We evaluate the proposed method using MMDetection[32], which is known for its state-of-the-art detection performance. The evaluation metrics used are the average precision (AP) and mean average precision (mAP) of object detection at 50%
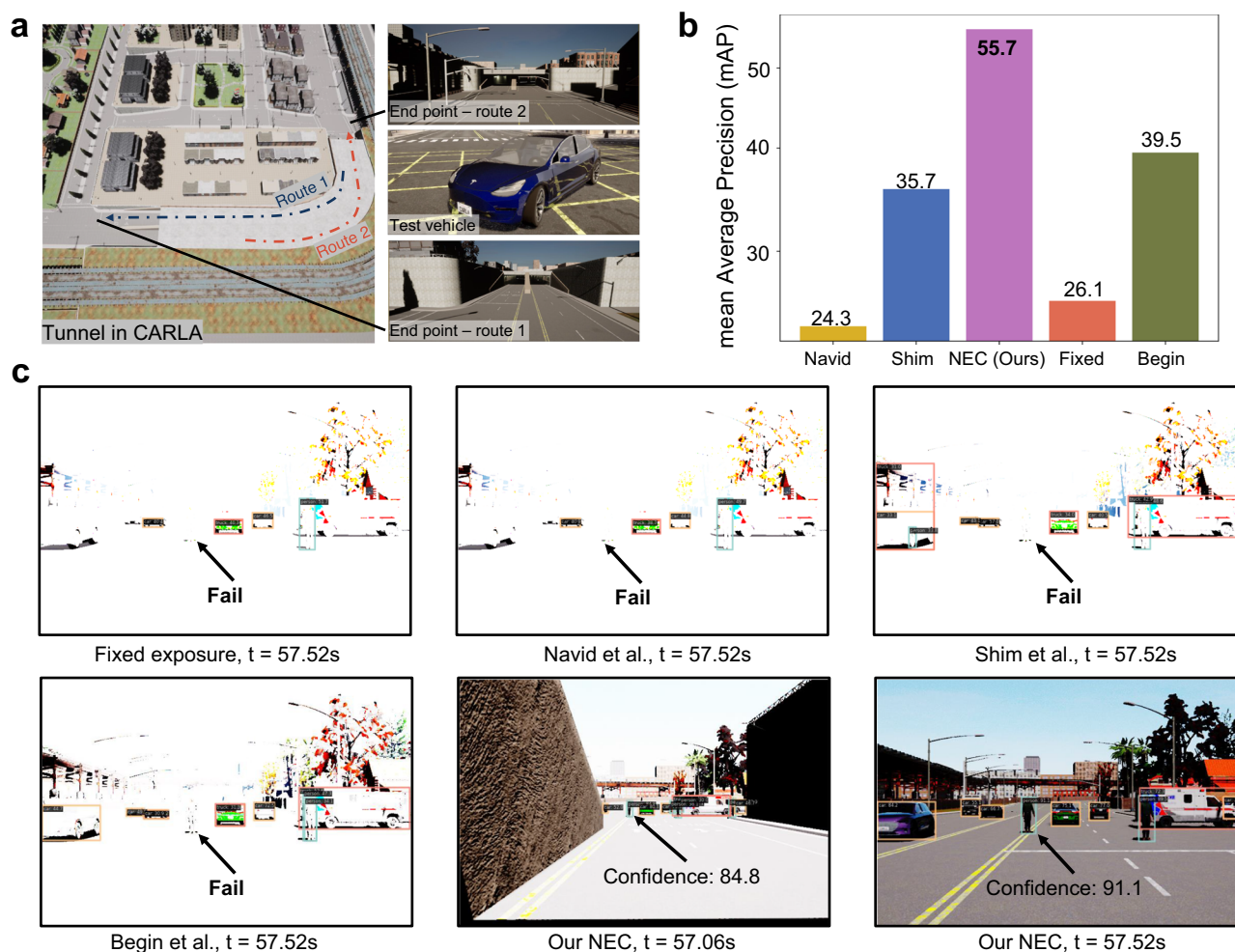
Fig. 2 | Evaluation on simulated driving. a The tunnel environment, test vehicle, and ending points of the two routes in the CARLA simulator. **b** Average detection precision using images generated by different exposure control methods. **c** Comparison for the first detection time of a specific pedestrian on route 1. The proposed NEC enables an early detection of the pedestrian, ≈460 ms (57.06 s versus 57.52 s), indicated by the back arrow. This early adjustment provides the vehicle with nearly 26 m for emergency braking.

intersection over union (IoU), as defined in the Supplementary Section 4.1. The comparison in Fig. 2b illustrates the performance of our NEC with Shim[13], Navid[10], Begin[18], and fixed exposure. Across all three classes (car, person, and truck) involved in the synthetic evaluation, our NEC outperforms others by a large margin. Regarding the overall mAP performance, the NEC outperforms Shim by 56% (55.7 versus 35.7), Navid by 129% (55.7 versus 24.3), and Begin by 14% (55.7 versus 39.5). This is attributed to the NEC's ability to rapidly adjust camera parameters (exposure time in Fig. 3a, gain in Fig. 3b) when the vehicle exits the tunnel. Consequently, the system effectively reduces the over-exposed rate below a valid threshold (50%, Fig. 3c).

In comparison, other benchmark methods nearly reach full saturation (Fig. 3c, dotted line, fully saturated). This occurs primarily because conventional exposure control relies solely on the image sensor for control. As the vehicle travels at a high speed (221 km h^{-1}), the natural lighting drastically increases upon exiting the tunnel (Fig. 3d, 57.06 s). At this moment, Navid solely utilizes histogram information for feedback adjustment, resulting in slow adjustment and pronounced over-exposure (Fig. 3d, 57.09 s), similar to manual EC. Shim employs image gradient prediction, yielding slightly better performance than Navid (35.7 mAP for Shim, 24.3 mAP for Navid). However, under such intense dynamic lighting conditions, the image itself may become saturated, removing the essential gradient information required by Shim and increasing its latency (Fig. 3d, Shim, 70.76 s).

Begin[18] fuses the optical flow with the prediction of saturation, thus achieving faster adjustment speed and slightly better performance (39.5 mAP for Begin) compared with Shim (35.7 mAP). In contrast, the NEC, by design, is highly sensitive to minor irradiance changes and provides a low-latency control signal to the camera. Consequently, the NEC can properly adjust the camera immediately after the illumination change and well before image capture commences, effectively mitigating over-exposure (Fig. 3d, 57.09 s). For more details, please refer to Supplementary Movie 1.

Emergency Response. The detection of pedestrians is of vital importance in autonomous or assisted-driving vehicles, particularly in scenarios where on-vehicle cameras encounter saturation problems that can greatly delay the response time of the decision system. In Fig. 2c, we present evidence that demonstrates the successful detection of a pedestrian in the middle of the road by the on-vehicle camera using the proposed NEC, with an 84.8% confidence level at time ≈460 ms earlier than conventional EC methods. Considering the vehicle's average speed of 200 km h^{-1}, such a reduction in latency provides the vehicle with nearly 26 meters of additional safe braking distance if needed.

Real-world evaluation. We further validate the proposed NEC in real-world vehicles (Fig. 4a). The driving scenarios are typical highway tunnels on a sunny day with minor clouds. The system setup and
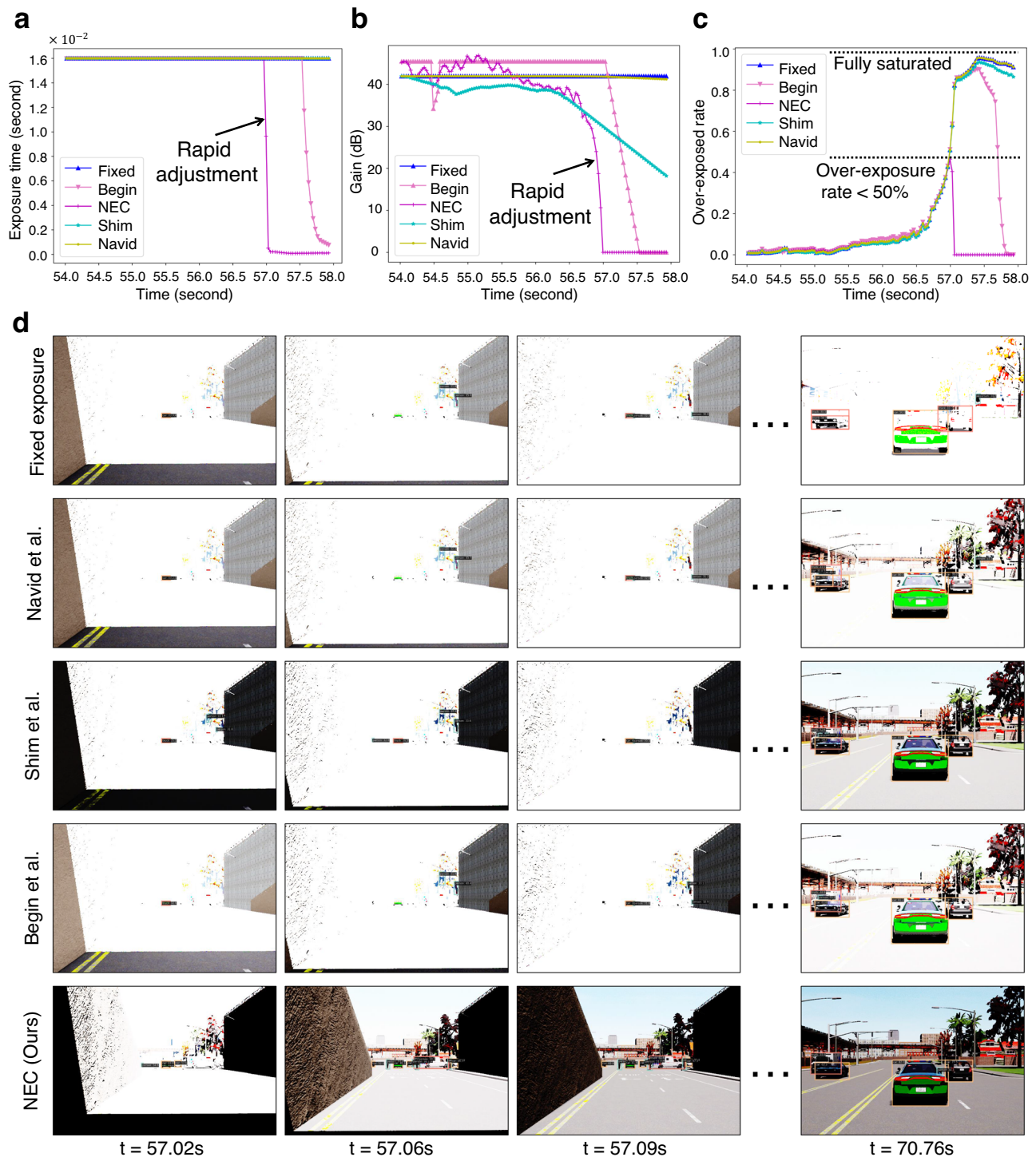
Fig. 3 | Comparison of the proposed NEC and baselines (Shim[13], Navid[10], Begin[18], and fixed exposure) for object detection at the exit of the tunnel on route 1. a Exposure time. b Gain. c Over-exposed rate. d Three consecutive frames from 57.02 s to 57.09 s demonstrate the rapid adjustment capability of our NEC method. In contrast, the conventional methods (Shim[13], Navid[10], and Begin[18]) take a significant amount of time to adapt from the dark indoor lighting to the bright sunlight. This convergence process often takes dozens of seconds (70.76 s).

driving details are provided in Supplementary Section 4.3. In Fig. 4d, the vehicle is facing direct sunlight (when exiting the tunnel at $t \approx 137.5$ s), which rapidly increases the illumination. At this moment, conventional solutions (Shim and Navid) fail to rapidly adjust the exposure setting (Fig. 4b), resulting in an intense increase of the over-exposed rate (Fig. 4d). In Fig. 4c, we observe that while Shim starts to reduce the gain nearly at the same time as NEC, it still cannot predict a physically correct exposure time (Fig. 4b, the zoom-in area), leading to the saturated image and failing the object detection as shown in Fig. 4f, $t = 137.5$ s to $t = 140$ s. The proposed NEC, in comparison, can compute physically correct parameters and directly set optimal exposure parameters, alleviating the saturation problem (Fig. 4f, $t = 137.5$ s) and successfully maintaining a much lower saturation rate (over-exposed rate of 40% in Fig. 4d) to ensure success and continuous object
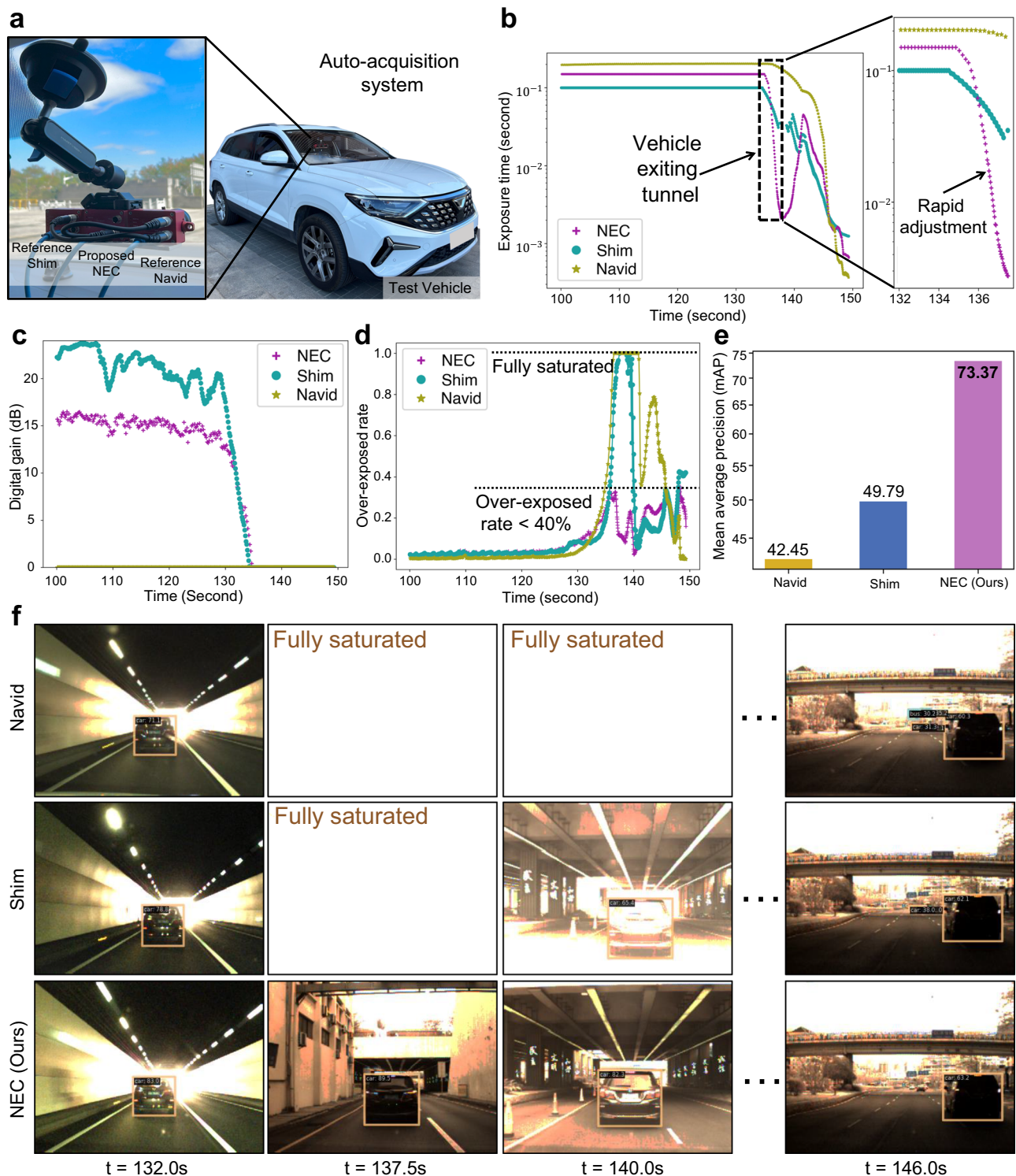
Fig. 4 | Evaluation on real-world driving. a Acquisition system. **b** Exposure time. **c** Gain. **d** Over-exposed rate. **e** Mean average precision (mAP). **f** Object detection results when the acquisition vehicle is driving out of a tunnel. The proposed NEC effectively manages the substantial increase in natural lighting when the vehicle exits the tunnel, ensuring consistent detection by swiftly adapting to this drastic change in lighting conditions and preventing the occurrence of over-exposed images. This avoids any potential compromise in the accuracy of the detection process (see $t = 137.5$ s in **f**).

detection. In Fig. 4e, we observe a notable improvement in mAP with the proposed NEC (mAP 73.37), which is 72.8% better than Navid (mAP 42.45) and 47.3% better than Shim (mAP 49.79). Refer to the Supplementary Movie 2 for full details.

To evaluate the robustness of the proposed NEC, we provide a long-distance driving evaluation (7.4 km) in Supplementary

Section 4.8, Supplementary Movie 8, and 9. We compare the proposed NEC with Shim and Begin in the experiment, directly showing that the proposed NEC can robustly function in daily scenes with performance similar to conventional exposure controls while enabling much faster adjustment in challenging lighting conditions (e.g., the tunnel exit).
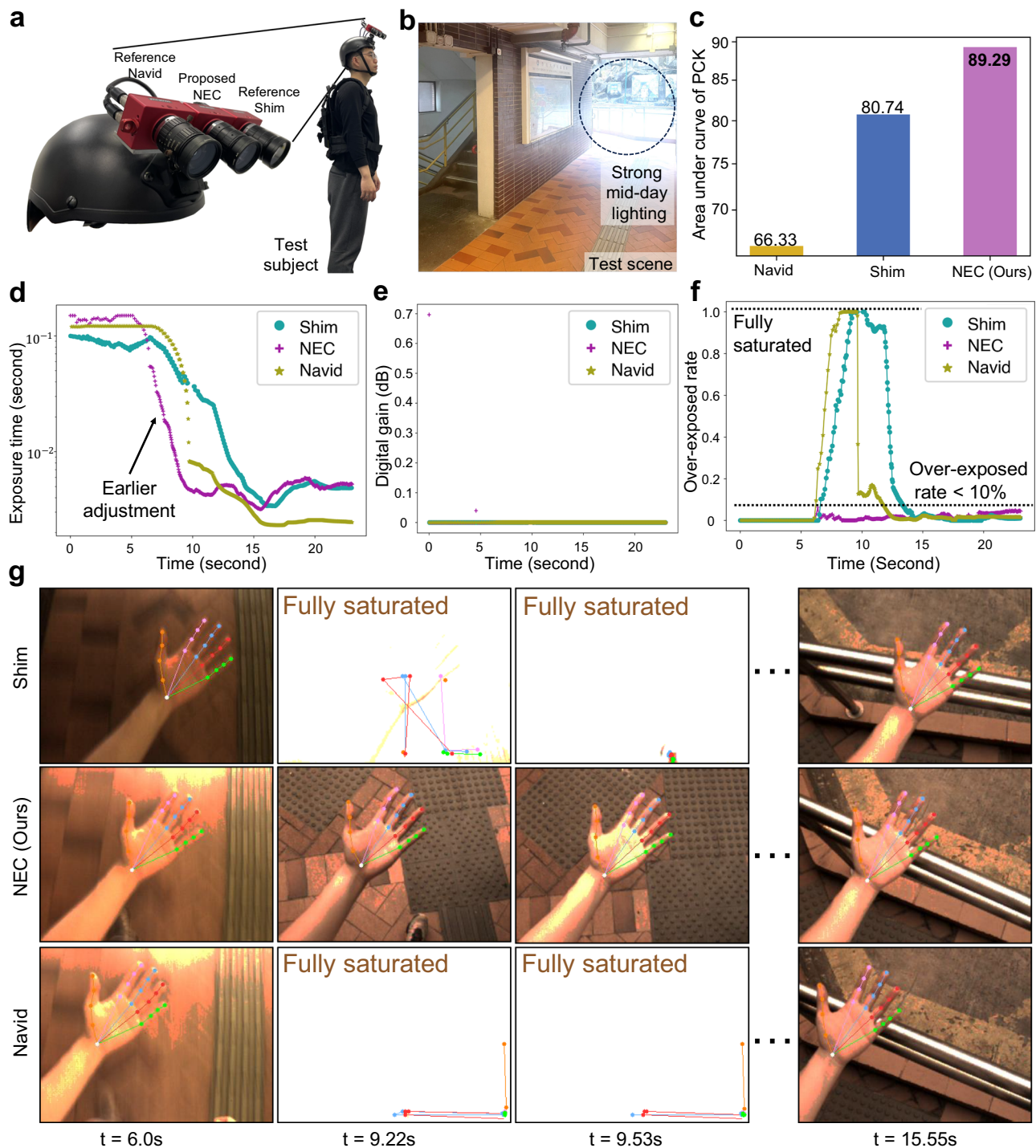
Fig. 5 | Evaluation on hand-pose detection for mixed reality. a Head-mounted hand pose recognition system. **b** Test scene. **c** Area under the curve of the percentage of correct keypoints (PCK). **d** Exposure Time. **e** Gain. **f** Over-exposed Rate. **g** Detection results of hand pose recognition. The proposed NEC effectively manages the significant increase in natural lighting when the test subject walks out of the building, ensuring consistent and reliable detection of the user's hand (see $t = 9.22$ s in **g**).

In-field hand-pose detection for mixed reality

Detecting hand pose in the field is essential but challenging[38], particularly when encountering the in-field usage of mixed reality (MR) devices, where users frequently transition between indoor and outdoor environments. These transitions often come with strong lighting variations that conventional exposure control is unable to handle, resulting in longer latency and a negative impact on the user experience.

In Fig. 5a, we utilize a head-mounted device equipped with three downward-facing DAVIS cameras to capture human hand movements. The testing environment involves typical indoor-outdoor transitions in common urban environments (Fig. 5b, more details in Supplementary Section 4.4). In Fig. 5d-g, the test subject walks out of a building and encounters a sudden increase in overall illumination due to direct sunlight (Fig. 5g, $t = 9.22$ s). This increase in illumination leads to strong saturation in the images captured by Navid and Shim, resulting in an

intense rise in the over-exposed rate (Fig. 5f). As depicted in Fig. 5g, image-based exposure control can result in the over-exposure problem, causing state-of-the-art hand pose detection (RTMPose[33]) to perform poorly on these saturated images. Consequently, the overall performance percentage of correct keypoints (PCK) decreases (Shim: 80.74, Navid: 66.33 in Fig. 5c). With the proposed NEC, the transition from indoor to outdoor is smooth (due to early adjustment in Fig. 5d), and the over-exposed rate remains consistently low (below 10% in Fig. 5f), which benefits keypoint detection, allowing it to consistently function and maintain a high PCK score (89.29 in Fig. 5c). Refer to the Supplementary Movie 3 for more details.

The proposed NEC could also benefit medical AR/VR devices in surgical lights. In Supplementary Section 4.7, we provide an evaluation of the proposed NEC using professional treatment luminaires in a dental office. The results directly show that the proposed NEC enables earlier adjustments and successfully removes saturation during diagnosis. This not only reduces display latency but also improves the detection accuracy of human facial keypoints, providing a better user experience for the doctor. We provide Movie description in Supplementary Movie 6 and 7.

Feature detection for finer 3D reconstruction

Vision-based 3D mapping, or simultaneous localization and mapping (SLAM), serves as a fundamental module for robotics[39] as well as spatial computing devices[40]. However, in real-world environments, rapid illumination changes pose significant challenges to modern machine vision sensors, resulting in over/under-exposure problems. Although these problems may occur briefly and only affect a few images, they can significantly disrupt the continuous feature tracking necessary for vision-based trajectory estimation. As a result, this disruption can cause significant deviations in the trajectory and ultimately have a detrimental impact on the quality of the 3D reconstruction.

In Fig. 6, we utilize a handheld 3D scanning device equipped with three cameras working simultaneously to scan historic buildings during midday (more details in Supplementary Section 4.5 and Supplementary Movie 4). The scanning covers the surface of the structures following the trajectory illustrated in Fig. 6a. During midday, the structure surface presents a mixture of complex lighting conditions that require exposure control to quickly adjust its parameters during scanning. However, conventional EC algorithms are slow and often result in over/under-exposure (Fig. 6f, under-exposure at $t = 16.37$ s and over-exposure at $t = 24.99$ s for Shim and Navid, respectively), which significantly reduces the matched inliers of the scale-invariant feature transform (SIFT[34]). These problems happened across the whole scanning (Fig. 6e, both Shim and Navid easily reach full saturation). With rapid adjustment (e.g., Fig. 6d, ~1 s earlier), the over-exposure rate of the proposed NEC rarely exceeds 35%.

Figure 7 presents the 3D reconstruction results using images captured by Navid[10], Shim[13], and our NEC. We observe that the models in Fig. 7a and b fail to reconstruct the bright side of the structure due to the tracking loss caused by the image over-exposure. In contrast, Fig. 7c presents a complete and fine model reconstructed from unsaturated images captured by our NEC.

Runtime evaluation

We evaluate the computation efficiency of the proposed NEC in event processing rate and the real-time factor. The event processing rate indicates the number of events data the system can process with each second. The real-time factor is defined as the ratio of the total interval of the data (i.e., events and frames) to the total processing time. Therefore, achieving a real-time performance requires a factor >1. Since the major bottleneck lies in the event processing for the double integral, we compare the proposed P-TEDI with three other state-of-

the-art asynchronous event double integrals, including our Fast-EDI[41], EDI[42], and mEDI[26]. To ensure a fair comparison, the benchmark is performed on the same computation platform using a single central processing unit (CPU) of a laptop equipped with the Intel i7-10870H@2.2 GHz and 32 GB random access memory (RAM). All methods are implemented using C++. Table 1 presents the evaluation results. Overall, the proposed P-TEDI is 8 × faster than Fast-EDI[41] and 623 × faster than mEDI, demonstrating the efficiency improvement of the proposed peripheral conversion. Regarding the real-time factor, the proposed P-TEDI achieves a value of 19, which is significantly >1, suggesting a real-time running performance is viable even on low-end devices.

Discussion

In this work, driven by the bio-principles from human visual systems, i.e., the pupillary control pathway and peripheral vision, we have designed and implemented the neuromorphic exposure control (NEC), which fundamentally alleviates the saturation problem that long troubling real-world intelligent systems. The NEC, trilinear event double integral, and peripheral event/image conversions break the circular dependency in previous research and present a new paradigm for the exposure control problem. The high efficiency, low latency, and bio-inspired nature of NEC provide timely and robust neuromorphic synergy for lighting-robust machine vision at a low cost, benefiting a substantial number of real-world intelligent applications and guiding the design of next-generation exposure control in machine vision.

Limitations

The proposed NEC has two major limitations. On the hardware side, the DAVIS camera does not support re-capture during exposure. Therefore, if challenging lighting changes occur during the exposure time, one frame will inevitably be over-/under-exposed. Our workaround is to use the direct integration from the previous frame to skip this frame in the TEDI framework to ensure robustness. The better solution is to re-capture the frame immediately at the hardware level and this could be achieved by redesigning the ISP in future work.

On the algorithmic side, the proposed NEC is purely based on the computation of physical lighting and is not specifically optimized for each downstream task. In future works, it is possible to add metrics like gradient or entropy into the pipeline to evaluate the video stream and optimize the exposure parameters to improve performance for each downstream task.

This work has focused on employing the DAVIS camera to validate the principle of NEC. Moving forward, we envision exploring additional opportunities for tailored sensor configurations, such as RGB-event, infrared-event, and HDR-event. This exploration raises multiple novel research inquiries, including spatial-temporal calibration, optical/mechanical design, fusion algorithms development, optimal synchronization, and integration with the camera driver system.

Methods

Control framework

The primary objective of NEC is to reduce the saturation in image capturing by rapidly setting appropriate exposure parameters. To this end, we leverage both events and frames to estimate scene irradiance, allowing us to determine the optimal exposure values, including exposure time and gain.

Exposure time. We utilize the log inverse camera response function (log-iCRF) to relate the irradiance $L(\mathbf{x})$, exposure time T, and image intensity $I(\mathbf{x})$, with $\mathbf{x} = (x, y)$ denoting the pixel position. The camera response function[25] is given by $I(\mathbf{x}) = f_{CRF}(H(\mathbf{x}))$, where $H(\mathbf{x})$ is the exposure of pixel $\mathbf{x}$ given by $H(\mathbf{x}) = T \cdot L(\mathbf{x})$. Then, the log-iCRF can be
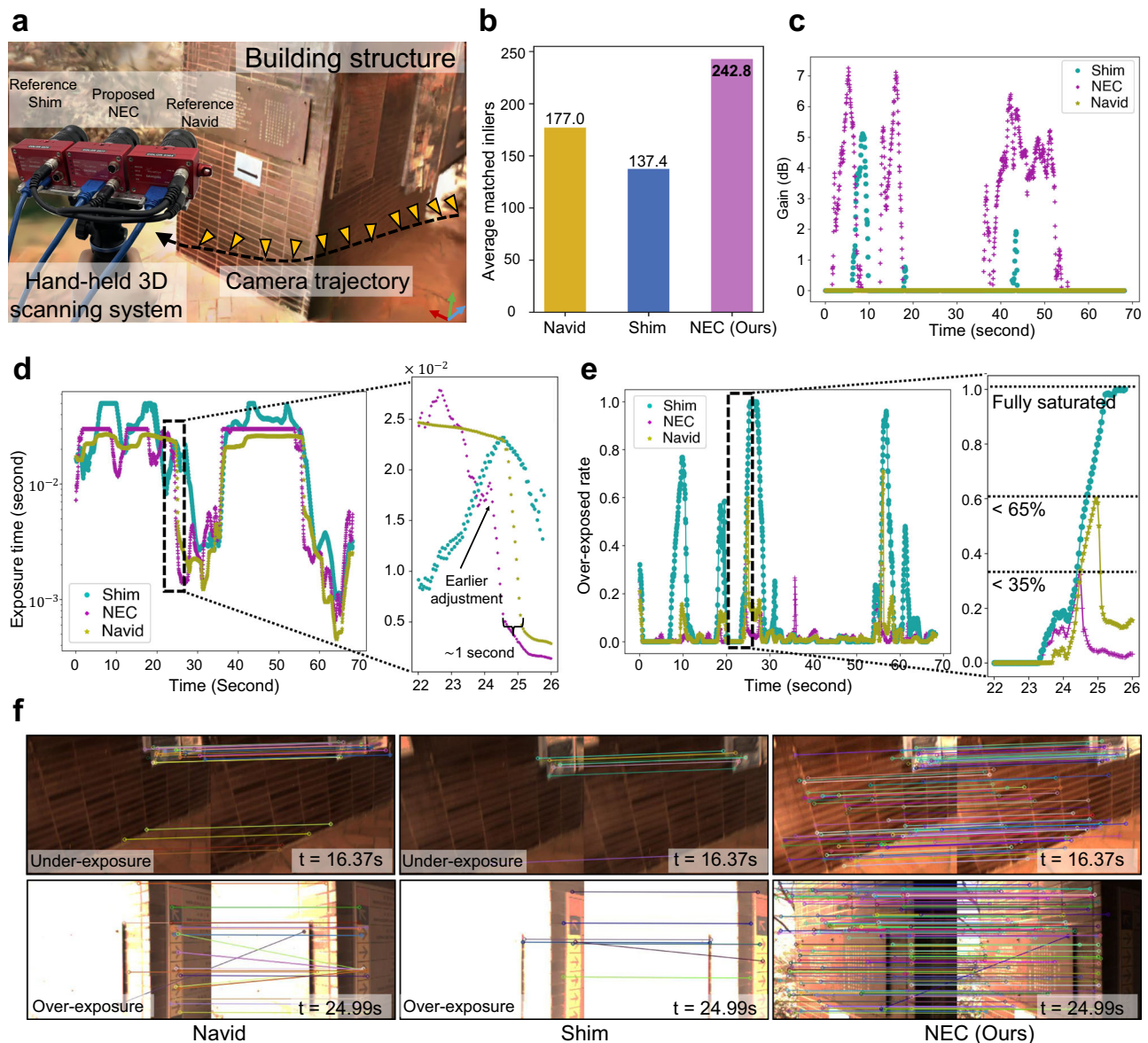
Fig. 6 | Evaluation on 3D reconstruction. a Hand-held 3D scanning system for scene reconstruction. **b** Matched inliers of the scale-invariant feature transform (SIFT) features[34]. **c** Gain. **d** Exposure time. **e** Over-exposed rate. **f** Feature matching of SIFT. The over/under-exposure problems significantly impact the accuracy of matches (see **f**, $t = 16.37$ s and $t = 24.99$ s, as evidenced by Navid and Shim), and disrupt the continuous feature matching, which is substantial for visual localization and 3D reconstruction. The proposed NEC yields unsaturated, high-fidelity, substantially increased accurate matches (242.85 in **b**).

represented as:

$$\ln f_{CRF}^{-1}(I(\mathbf{x})) = \ln T + \ln L(\mathbf{x}), \qquad (1)$$

with $\ln(\cdot)$ denotes the natural logarithm operator. As shown in Fig. 8d, the log-iCRF function allows us to input the desired intensity value and obtain the desired underlying exposure. Therefore, we can compute an optimal intensity value to determine the optimal exposure and avoid saturation by utilizing the log-iCRF and scene irradiance $L(\mathbf{x})$.

To avoid over/under-exposure, we need to place the pixel intensity in the middle of the dynamic range to increase the separation between two edges, as shown in Fig. 8d. However, for off-the-shelf APS sensors, the exposure time setting is applied uniformly to all pixels. Therefore, we require a unified intensity value applied to all pixels. To achieve this, we utilize the log-space mean irradiance to derive an intensity value. By setting the mean value $mean(\ln L(t))$ to the middle of the dynamic range, we effectively position the majority of pixels at

this optimal, thereby avoiding the saturation problem. The optimal exposure time T_{opt} can be obtained as:

$$T_{opt} = \exp\left(\ln f_{CRF}^{-1}(I_{opt}) - mean(\ln L(t))\right), \qquad (2)$$

where $I_{opt} = f_{CRF}(\exp(\frac{g(250) + g(9)}{2}))$ is the goal intensity that represents the middle of the dynamic range. We exclude a number of nonlinear digits near the edge. In this case, the valid digits range from 9 to 250.

To accurately estimate the mean irradiance, we develop a new model, dubbed trilinear event double integral (discussed in "Physical relation for events and frames"). We have also adapted this model to the peripheral vision to enhance the computation speed (discussed in "Fast irradiance computation in peripheral vision"). This improvement greatly reduces the computation load, enabling real-time irradiance estimation for high-rate exposure control.
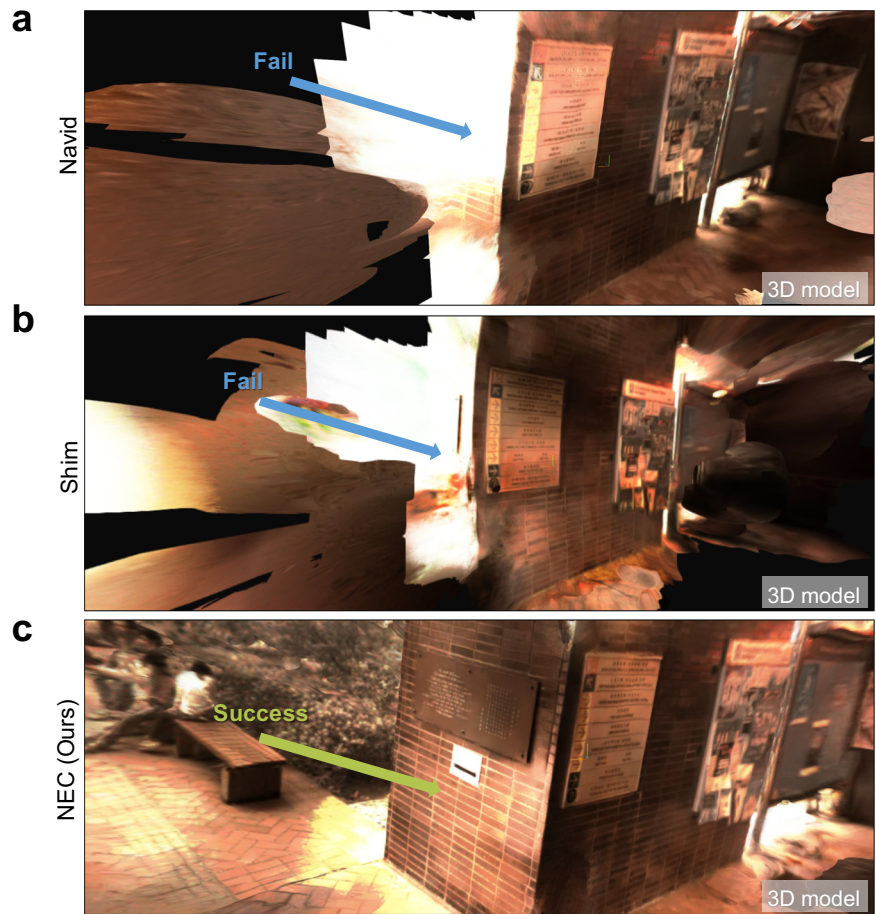
Fig. 7 | Evaluations of AEs in challenging the 3D reconstruction scene. Historical building models reconstructed using images captured by (**a**) Navid[10]; (**b**) Shim[13]; (**c**) the proposed NEC. The proposed NEC allows for accurate and comprehensive 3D reconstruction, whereas the other methods fall short in achieving this level of detail.

The control framework updates the exposure time in the following manner:

$$T_{set} = T_i + f_w(T_{err})T_{err}, \quad (3)$$

where $T_{err} = (T_i - T_{opt})$ represents the error; T_{set} is the exposure time we seek to set for the next image capturing; T_i is the exposure time of i image; $f_w(\cdot)$ is a weighting function (Fig. 8e) to weight the error, reducing the flashing effect when the change in irradiance is subtle.

Gain. The exposure parameter is represented using both time and gain. Once the time T_{set} exceeds the maximum bound, the gain is then adjusted. This strategy adjusts the exposure time first to minimize the

Table 1 | Runtime comparisons against existing solutions

Method	Event Processing Rate (↑) (Million Events per Second)	Real-time Factor (↑) (Real-time ≥ 1)
EDI[42]	1.11 ± 0.0036	0.47 ± 0.01
mEDI[26]	0.21 ± 0.0002	0.081 ± 0.00017
Fast-EDI[41]	15.84 ± 0.1	2.03 ± 0.014
P-TEDI (Proposed)	130.82 ± 2.17	19.01 ± 0.3

↑ means the higher the value the better. We provide the event processing rate and real-time factor of three different algorithms with standard error, tested under identical conditions. The real-time factor is defined by dividing the total runtime by the event duration, real-time performance is achieved if the factor is >1.
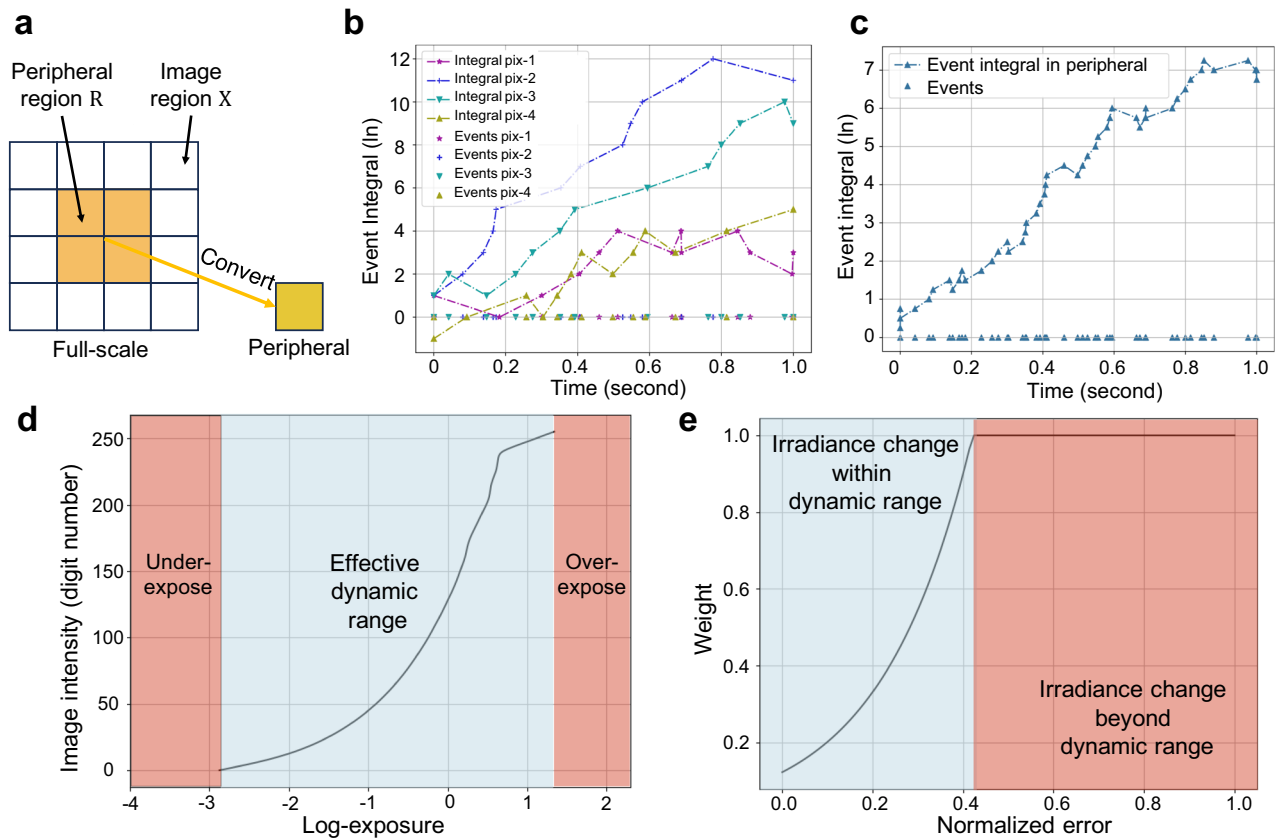
decrease of signal-to-noise ratio caused by increasing the gain[17]. The gain K_{i+1} and exposure time T_{i+1} of the $i + 1$ image are calculated as follows:

$$K_{i+1} = max(1, \frac{T_{set}}{T_{max}}), \quad T_{i+1} = \frac{T_{set}}{K_{i+1}}. \quad (4)$$

In each update of the mean irradiance, we compute the corresponding gain and exposure time using Equation (2) to (4). These computed values are then sent to the camera, which adjusts its parameters for image acquisition accordingly. Once the camera is ready, it captures an image without over/under-saturation. It is important to highlight that our method relies solely on the physical computation of irradiance and does not make any assumptions about the scenes being captured. This makes our approach more general and robust compared to methods that rely on explicit assumptions such as object consistency[17] or image gradient consistency[13].

Physical relation for events and frames

The proposed NEC leverages physical estimation of the scene irradiance to support its high-rate update of the control parameter. However, accurately estimating the scene irradiance presents a significant challenge. Directly accumulating events can be biased by noise, and the presence of motion blur can significantly impact the accuracy of irradiance estimation from images. Therefore, we propose a joint event-frame integration model to compute the irradiance signal. Conventionally, this is achieved by the event double integral (EDI)[26]. However, EDI suffers from inaccuracies and strong non-linearity in

Fig. 8 | Principle of conversion and control. Conversion of (**a**) an image and (**b**, **c**) events to the peripheral vision. To adapt image data to the peripheral vision, we seek to determine the average image intensity. For event data, we reconstruct the irradiance-change curve of each event and employ logarithmic space to generate a new mean curve, maintaining the original data's event count. **d** Camera response function of the DAVIS346 camera. **e** Weight function for exposure time adjustment.

threshold parameters when facing strong lighting variations (refer to Supplementary Section 1.1 for detailed experimental analysis), leading to time-consuming optimizations for parameter identification[26,43].

Trilinear Event Double Integral (TEDI). By modeling the temporal formation process of the event camera using linear parameters, we seek to establish the connection between events and frames (refer to Supplementary Section 1.2.3 for complete derivation), enabling efficient online irradiance prediction. The complete form for irradiance computation is derived by:

$$\ln L(g, \mathbf{x}) = \ln \bar{L}(\mathbf{x}) - \ln E_{\text{tri}}(g, \mathcal{T}, \mathbf{x}), \text{ with}$$
$$E_{\text{tri}}(g, \mathcal{T}, \mathbf{x}) = \frac{1}{|\mathcal{T}|} \int_{t \in \mathcal{T}} \exp(\mathbf{c}^{\top} \mathbf{D}(g, t, \mathbf{x})) dt, \quad (5)$$

where $L(g, \mathbf{x})$ is the irradiance of pixel $\mathbf{x}$ at time g; $\bar{L}(\mathbf{x})$ represents the temporal average irradiance obtained from the image $I(\mathbf{x})$ using the CRF, i.e., $\bar{L}(\mathbf{x}) = \frac{1}{|\mathcal{T}|} f_{\text{CRF}}^{-1}(I(\mathbf{x}))$; $\mathcal{T}$ is the exposure interval for the image $I(\mathbf{x})$. $\mathbf{c}$ defines the positive and negative threshold values, i.e., c^+ and c^-; $\mathbf{D}(g, t, \mathbf{x})$ defines the direct integral; these two vectors are given by:

$$\mathbf{c} = \begin{bmatrix} c^+ \\ c^- \end{bmatrix}, \mathbf{D}(g, t, \mathbf{x}) = \begin{bmatrix} D^+(g, t, \mathbf{x}) + D_c^+(g, t, \mathbf{x}) \\ D^-(g, t, \mathbf{x}) + D_c^-(g, t, \mathbf{x}) \end{bmatrix}, \quad (6)$$

where $D^+(g, t, \mathbf{x})$ and $D^-(g, t, \mathbf{x})$ are the direct integral for positive and negative events, respectively; $D_c^+(g, t, \mathbf{x})$ and $D_c^-(g, t, \mathbf{x})$ represent the linear compensation defined by linear parameters $\{a^+, b^+, a^-, b^-\}$ (refer to Supplementary Section 4.2 for detail formulations). The compensation is applied to the direct integral once the linear function yields a positive value. Therefore, for smooth motion and lighting, the modeling is the same as the conventional model, consistent with the theory

threshold obtained from the physical settings[44]. For radical motion or lighting changes, the compensation will reduce the prediction error.

Optimization. The proposed TEDI is a model-based method that requires calibration of the parameters. Once done, the model can provide an accurate prediction of the irradiance using events and frames.

We calibrate the overall 6 parameters i.e., $\mathbf{p} = \{a^+, b^+, c^+, a^-, b^-, c^-\}$ using the least square solver[45]. Given two images and corresponding events, we utilize the TEDI to recover the latent irradiance at the same moment, e.g., $L_i(g, \mathbf{x}) = L_{i-1}(g, \mathbf{x})$, where $L_i(g, \mathbf{x})$ is the latent irradiance reconstructed using image I_i at time g, $L_{i-1}(g, \mathbf{x})$ using image I_{i-1}. Therefore, using a video sequence with N_{img} images and corresponding events, we can establish the irradiance consistency across frames. The optimal parameters $\mathbf{p}^*$ can be estimated by solving the optimization problem:

$$\mathbf{p}^* = argmin_{\mathbf{p}} \sum_{i=1}^{N_{\text{img}}} \sum_{\mathbf{x} \in \mathbf{X}} L_i(g, \mathbf{x}) - L_{i-1}(g, \mathbf{x}). \quad (7)$$

We use the theory threshold[44] for initialization of the c^+ and c^+ and optimize all linear parameters (results in Supplementary Table S1). These linear parameters reduce the prediction error for large lighting variations (see Supplementary Fig. S2h for comparisons), making possible accurate prediction of the irradiance.

Fast irradiance computation in peripheral vision
As discussed in Equation (2), determining the optimal exposure time relies on calculating the mean irradiance by averaging the irradiance

$L(g, \mathbf{x})$ in Equation (5). However, the computational burden of computing the double integral in such scenarios is overwhelming, particularly with event cameras that generate millions of events per second. While our TEDI approach utilizes linear parameters to avoid online parameter optimization, processing millions of events per second remains highly challenging. Previous work[41] demonstrated that buffering the exponential results of the event integral can achieve linear processing time for the double integral model. However, the Fast-EDI method[41] requires extensive memory buffering, with each event generating a new computation result. Storing millions of these results necessitates frequent access to various memory blocks, ultimately limiting the method's overall performance.

The human visual system employs a fascinating bio-principle that reduces the number of photoreceptor cells in the peripheral region to obtain a wider general view. This conversion optimizes the efficiency of bio-information processing in the brain. Motivated by the efficient processing strategy of peripheral vision in the human visual system, we propose the Peripheral-TEDI, a highly efficient processing paradigm that significantly alleviates the processing burden.

We first define the peripheral vision that converts pixels in the peripheral region R of the image plane to a single pixel representing the overall lighting condition of the region. For this process, we utilize the log scale mean of the irradiance for conversion, giving as:

$$\ln P(t) \approx \frac{1}{N_R} \sum_{\mathbf{x} \in \mathbf{R}} \ln L(t, \mathbf{x}), \tag{8}$$

where $P(t)$ is the irradiance in the peripheral region, N_R is the number of the peripheral region R. Similar to the formation process of an image (detailed in Supplementary Section 1.1.2), the exposure in the peripheral region equals the integration of irradiance over exposure time, which also equals the temporal average irradiance $\bar{P}$ times the exposure time:

$$|\mathcal{T}|\bar{P} = \int_{t \in \mathcal{T}} P(t)dt. \tag{9}$$

For the image data, as illustrated in Fig. 8a, we take the mean of the intensity of pixels in the region R to approximate, indicated as:

$$|\mathcal{T}|\bar{P} \approx f_{CRF}^{-1}\left(\frac{1}{N_R} \sum_{\mathbf{x} \in \mathbf{R}} I(x)\right). \tag{10}$$

For the events, we take the mean of the log space event direct integral to generate a new irradiance curve in the peripheral region, indicated as:

$$\mathbf{D}_p(g, t) = \frac{1}{N_R} \sum_{\mathbf{x} \in \mathbf{R}} \mathbf{D}(g, t, \mathbf{x}), \tag{11}$$

where $\mathbf{D}_p(g, t)$ is the direct integral in the peripheral region. This process reduces the demand for accessing different positions $\mathbf{x}$ but necessarily preserves all events for high accuracy. In the peripheral region (Fig. 8c), the coordinate is reduced to one and the direct integral is the mean of all integrals (e.g., integrals 1 to 4 in Fig. 8b). Therefore, the whole process equates all pixels in the image plane to one large pixel, and the event of each previous small pixel is equivalent to the event in the large pixel with a smaller contribution to illumination.

After the conversion, the forward prediction in the peripheral region is given by:

$$P(t) = P(g) \exp(\mathbf{c}^{\top}\mathbf{D}_p(g, t)). \tag{12}$$

Then, incorporating Equation (12) into Equation (9) yields a double integral in the peripheral vision (refer to the Supplementary Section 3

for the complete derivation):

$$T\bar{P} = P(g)\int_{t \in \mathcal{T}} \exp(\mathbf{c}^{\top}\mathbf{D}_p(g, t))dt. \tag{13}$$

With taking the log of both sides, the TEDI in the peripheral vision (Peripheral-TEDI) is expressed as:

$$\ln P(g) = \ln \bar{P} + \ln E_{p-tri}(g, \mathcal{T}), \quad \text{with}$$
$$E_{p-tri}(g, \mathcal{T}) = \frac{1}{|\mathcal{T}|}\int_{t \in \mathcal{T}} \exp(\mathbf{c}^{\top}\mathbf{D}_p(g, t))dt, \tag{14}$$

where $\ln \bar{P}$ is given by taking the log of both sides of Equation (10):

$$\ln \bar{P} \approx \ln f_{CRF}^{-1}\left(\frac{1}{N_R} \sum_{\mathbf{x} \in \mathbf{R}} I(x)\right) - \ln |\mathcal{T}|. \tag{15}$$

We set the peripheral region R as the entire image plane to estimate the general illumination for the exposure control system in Equation (2), i.e., $mean(\ln L(t)) \approx \ln P(t)$. By reducing the computation of TEDI to a single pixel using Peripheral-TEDI, we can significantly reduce the memory consumption and access time[41], while boosting the processing speed. In the Supplementary Section 3, we present comprehensive proof of the conversion from TEDI to P-TEDI for computing the mean irradiance, showing that P-TEDI can significantly improve efficiency without sacrificing accuracy.

Informed consent from the human research participant for the publication of their images have been obtained.

Reporting summary
Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability
The data used in this study are available in the repository [https://doi.org/10.25442/hku.26211053.v1].

Code availability
The source codes of this study are available in the repository [https://doi.org/10.25442/hku.26206397.v1]. This code is freely accessible. We provide the real-time executable version written in C++. All the information including system requirements, installation guide, updates, and instructions for use can be found on the code page.

References
1. Goddard, M. A. et al. A global horizon scan of the future impacts of robotics and autonomous systems on urban ecosystems. *Nat. Ecol. Evol.* **5**, 219–230 (2021).
2. Long, Y. et al. Relational graph learning on visual and kinematics embeddings for accurate gesture recognition in robotic surgery. In *2021 IEEE Int. Conference on Robotics and Automation (ICRA)*, 13346–13353 (IEEE, 2021).
3. O'Callaghan, J. *Apple Vision Pro: What Does it Mean for Scientists?* https://www.nature.com/articles/d41586-024-00387-z (2024).
4. Advisory, M. I. R. *Cmos Image Sensor Market Size & Share Analysis —Growth Trends & Forecasts (2024 - 2029)*. https://www.mordorintelligence.com (2024).
5. Mur-Artal, R., Montiel, J. M. M. & Tardos, J. D. Orb-slam: a versatile and accurate monocular slam system. *IEEE Trans. Robotics* **31**, 1147–1163 (2015).
6. Zhang, Z., Forster, C. & Scaramuzza, D. Active exposure control for robust visual odometry in hdr environments. In *IEEE International Conference on Robotics and Automation (ICRA)*, 3894–3901 (IEEE, 2017).

7. Brad, T. *Tesla in Taiwan Crashes Directly into Overturned Truck, Ignores Pedestrian, With Autopilot on.* https://www.forbes.com/sites/bradtempleton/2020/06/02/tes (2020).

8. Sampat, N., Venkataraman, S., Yeh, T. & Kremens, R. L. System implications of implementing auto-exposure on consumer digital cameras. In *Sensors, Cameras, and Applications for Digital Photography*. 100–107 (International Society for Optics and Photonics, 1999).

9. Kuno, T., Sugiura, H. & Matoba, N. A new automatic exposure system for digital still cameras. *IEEE Trans. Consum. Electron.* **44**, 192–199 (1998).

10. Nourani-Vatani, N. & Roberts, J. Automatic camera exposure control. In *Australasian Conference on Robotics and Automation 2007* 1–6 (Australian Robotics and Automation Association Inc., 2007).

11. Montalvo, M. et al. Acquisition of agronomic images with sufficient quality by automatic exposure time control and histogram matching. In *International Conference on Advanced Concepts for Intelligent Vision Systems* 37–48 (Springer, 2013).

12. Lu, H., Zhang, H., Yang, S. & Zheng, Z. Camera parameters auto-adjusting technique for robust robot vision. In *2010 IEEE International Conference on Robotics and Automation (ICRA)* 1518–1523 (IEEE, 2010).

13. Shim, I. et al. Gradient-based camera exposure control for outdoor mobile platforms. *IEEE Trans. Circuits Syst. Video Technol.* **29**, 1569–1583 (2018).

14. Tomasi, J., Wagstaff, B., Waslander, S. L. & Kelly, J. Learned camera gain and exposure control for improved visual feature detection and matching. *IEEE Robotics Automation Lett.* **6**, 2028–2035 (2021).

15. Kao, W.-C., Hsu, C.-C., Kao, C.-C. & Chen, S.-H. Adaptive exposure control and real-time image fusion for surveillance systems. In *2006 IEEE International Symposium on Circuits and Systems* 1–4 (IEEE, 2006).

16. Yang, M., Wu, Y., Crenshaw, J., Augustine, B. & Mareachen, R. Face detection for automatic exposure control in handheld camera. In *Fourth IEEE International Conference on Computer Vision Systems (ICVS)*, 17–17 (IEEE, 2006).

17. Onzon, E., Mannan, F. & Heide, F. Neural auto-exposure for high-dynamic range object detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7710–7720 (2021).

18. Bégin, M.-A. & Hunter, I. Auto-exposure algorithm for enhanced mobile robot localization in challenging light conditions. *Sensors* **22**, 835 (2022).

19. Dufaux, F., Le Callet, P., Mantiuk, R. & Mrak, M. *High Dynamic Range Video: From Acquisition, to Display and Applications* (Academic Press, 2016).

20. Kim, J. & Kim, M. H. Joint demosaicing and deghosting of time-varying exposures for single-shot hdr imaging. In *Proc. IEEE/CVF International Conference on Computer Vision*, 12292–12301 (IEEE, 2023).

21. Photography, I. Automatic control of exposure. Standard ISO 2721:2013 (International Organization for Standardization, 2013).

22. Belliveau, A. P., Somani, A. N. & Dossani, R. H. *Pupillary light reflex* (StatPearls Publishing, 2019).

23. Brandli, C., Berner, R., Yang, M., Liu, S.-C. & Delbruck, T. A 240 × 180 130 db 3 μs latency global shutter spatiotemporal vision sensor. *IEEE J. Solid State Circuits* **49**, 2333–2341 (2014).

24. Lichtsteiner, P., Posch, C. & Delbruck, T. A 128 × 128 120 db 15 μs latency asynchronous temporal contrast vision sensor. *IEEE journal of solid-state circuits* **43**, 566–576 (2008).

25. Grossberg, M. D. & Nayar, S. K. Modeling the space of camera response functions. *IEEE Trans. Pattern Anal. Mach. Intell.* **26**, 1272–1282 (2004).

26. Pan, L. et al. High frame rate video reconstruction based on an event camera. *IEEE Trans. Pattern Anal. Mach. Intell.* **44**, 2519–2533 (2020).

27. Ziemba, C. M. & Simoncelli, E. P. Opposing effects of selectivity and invariance in peripheral vision. *Nat. Commun.* **12**, 4597 (2021).

28. Polyak, S. L. *The Retina* (Univ. Chicago Press, 1941).

29. Weber, C. & Triesch, J. Implementations and implications of foveated vision. *Recent Patents Comput. Sci.* **2**, 75–85 (2009).

30. Bartolozzi, C., Indiveri, G. & Donati, E. Embodied neuromorphic intelligence. *Nat. Commun.* **13**, 1024 (2022).

31. Sandamirskaya, Y., Kaboli, M., Conradt, J. & Celikel, T. Neuromorphic computing hardware and neural architectures for robotics. *Sci. Robotics* **7**, eabl8419 (2022).

32. Chen, K. et al. MMDetection: Open mmlab detection toolbox and benchmark. *arXiv* https://doi.org/10.48550/arXiv.1906.07155 (2019).

33. Jiang, T. et al. Rtmpose: Real-time multi-person pose estimation based on mmpose. *arXiv* https://doi.org/10.48550/arXiv.2303.07399 (2023).

34. Lowe, D. G. Distinctive image features from scale-invariant keypoints. *Int. J. Comp. Vision* **60**, 91–110 (2004).

35. Kolekar, S., de Winter, J. & Abbink, D. Human-like driving behaviour emerges from a risk-based driver model. *Nat. Commun.* **11**, 4850 (2020).

36. Yu, W. et al. Online legal driving behavior monitoring for self-driving vehicles. *Nat. Commun.* **15**, 408 (2024).

37. Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A. & Koltun, V. Carla: An open urban driving simulator. In *Conference on robot learning*, 1–16 (PMLR, 2017).

38. Wen, F., Zhang, Z., He, T. & Lee, C. Ai enabled sign language recognition and vr space bidirectional communication using triboelectric smart glove. *Nat. Commun.* **12**, 5378 (2021).

39. Wu, H., Li, Y., Xu, W., Kong, F. & Zhang, F. Moving event detection from lidar point streams. *Nat. Commun.* **15**, 345 (2024).

40. Rogers, C. et al. A universal 3d imaging sensor on a silicon photonics platform. *Nature* **590**, 256–261 (2021).

41. Lin, S. et al. Fast event-based double integral for real-time robotics. *arXiv* https://doi.org/10.48550/arXiv.2305.05925 (2023).

42. Pan, L. et al. Bringing a blurry frame alive at high frame-rate with an event camera. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6820–6829 (IEEE, 2019).

43. Antil, H. & Sayre, D. Bilevel inverse problems in neuromorphic imaging. *Inverse Probl.* **39**, 094003 (2023).

44. Nozaki, Y. & Delbruck, T. Temperature and parasitic photocurrent effects in dynamic vision sensors. *IEEE Trans. Electron Devices* **64**, 3239–3245 (2017).

45. Agarwal, S., Mierle, K. & Team, T. C. S. *Ceres Solver.* https://www.ceres-solver.org/ (2023).

Author contributions

Shijie Lin conceived the idea and developed the theory. Yifan Peng and Jia Pan verified the analytical methods. Shijie Lin, Guangze Zheng, and Ruihua Han wrote the code. Shijie Lin, Guangze Zheng, and Ziwei Wang conceived and planned the experiments. Shijie Lin, Guangze Zheng, Ruihua Han, Zeqing Zhang, and Wanli Xing carried out the experiments. Shijie Lin, Guangze Zheng, and Ziwei Wang wrote the manuscript. Yifan Peng and Jia Pan supervised the project. All authors discussed the results and contributed to the final manuscript.

Competing interests

Additional information