INVITED PAPER

The Remarkable Visual Abilities of Nocturnal Insects: Neural Principles and Bioinspired Night-Vision Algorithms

Exceptionally competent nocturnal eyes offer a template for innovative night-vision video.

By Eric Warrant, Magnus Oskarsson, *Member IEEE*, and Henrik Malm

ABSTRACT | Despite their tiny eyes and brains, nocturnal insects have remarkable visual abilities. Recent work—particularly on fast-flying moths and bees and on ball-rolling dung beetles—has shown that nocturnal insects are able to distinguish colors, to detect faint movements, to learn visual landmarks, to orient to the faint pattern of polarized light produced by the moon, and to navigate using the stars. These impressive visual abilities are the result of exquisitely adapted eyes and visual systems, the product of millions of years of evolution. Even though we are only at the threshold of understanding the neural mechanisms responsible for reliable nocturnal vision, growing evidence suggests that the neural summation of photons in space and time is critically important: even though vision in dim light becomes necessarily coarser and slower, those details that are preserved are seen clearly. These benefits of spatio-temporal summation have obvious implications for dim-light video technologies. In addition to reviewing the visual adaptations of nocturnal insects, we here describe an algorithm inspired by nocturnal visual processing strategies—from amplification of primary image signals to optimized spatio-temporal summation to reduce noise—that dramatically increases the reliability of video collected in dim light, including the preservation of color.

KEYWORDS | Compound eye; denoising; image enhancement; insect; nocturnal vision; structure tensor; summation

I. INTRODUCTION

Every life form on our planet is the product of countless millions of years of biological evolution. Each species of plant, animal, and microorganism has been uniquely fashioned by the forces of natural selection, allowing them to become exquisitely adapted to the endless variety of habitats and lifestyles they collectively possess. This adaptation invariably results from the evolution of structural and physiological specializations that uniquely match plants and animals to the specific lives they lead, maximizing their chances of survival and reproduction. From a human perspective, these specializations often appear remarkable. Not only do they allow organisms to solve difficult physical problems, such as to withstand extreme heat or cold, to overcome the forces of gravity to fly, or to reliably detect the weak mating calls of a distant conspecific, but they are also highly energy efficient, as well as adaptable and robust, allowing organisms to cope with large deviations in normal environmental conditions in an automatic, low-cost, and self-organizing fashion. Such qualities are highly desirable in engineering design, where sensors and actuators should not only solve specific tasks well, but should also be robust against unexpected (and unpredictable) changes in the system. Such qualities are particularly

Manuscript received March 16, 2014; accepted May 30, 2014. Date of publication August 19, 2014; date of current version September 16, 2014. This work was supported by Toyota Motor Engineering & Development Europe, the U.S. Air Force Office of Scientific Research (AFOSR), and the Swedish Research Council.

E. Warrant is with the Department of Biology, Lund University, Lund SE-221 00, Sweden (e-mail: Eric.Warrant@biol.lu.se).
M. Oskarsson is with the Centre for Mathematical Sciences, Lund University, Lund SE-221 00, Sweden.
H. Malm is with Foster and Partners, London SW11 4AN, U.K.

Digital Object Identifier: 10.1109/JPROC.2014.2332533

desirable in robotics where for decades engineers have strived to create fully autonomous machines that can perform faultlessly in the absence of human intervention. Not surprisingly then, the quest to understand the biological principles that exquisitely adapt organisms to their environments and which allow them to solve complex problems, has led to unexpected insights into how similar problems can be solved technologically. These biological principles often turn out to be refreshingly simple, even "ingenious," seen from a human perspective, making them readily transferable to man-made devices. This mimicry of biology, known more broadly as "biomimetics," has already led to a great number of technological innovations, from Velcro (inspired by the adhesive properties of burdock seed capsules) to convection-based air-conditioning systems in skyscrapers (inspired by the nest ventilation system of the African fungus-growing termite). In this paper, we describe a recent biomimetic advance inspired by the visual systems of nocturnal insects.

Constituting the great majority of all known species of animal life on Earth, the insects have conquered almost every imaginable habitat (excluding oceans). And despite having eyes, brains, and nervous systems only a fraction the size of our own, insects rely on vision to find food, to navigate and home, to locate mates, escape predators, and migrate to new habitats. Even though insects do not see as sharply as we do, many experience many more colors, can see polarized light, and can clearly distinguish extremely rapid movements. Remarkably, this is even true of nocturnal insects, whose visual world at night is around 100 million times dimmer than that experienced by their day-active relatives. Research over the past 15–20 years—particularly on fast-flying and highly aerodynamic moths and bees and on ball-rolling dung beetles—indicates that nocturnal insects that rely on vision for the tasks of daily life invariably see extremely well (reviewed by Warrant and Dacke [81]). We now know, for instance, that nocturnal insects are able to distinguish colors [31], [66], to detect faint movements [70], to avoid obstacles during flight by analyzing optic flow patterns [3], to learn visual landmarks [57], [67], [77], to orient to the faint polarization pattern produced by the moon [12], and to navigate using the faint stripe of light provided by the Milky Way [11].

To see so well at night, these insects, whose small eyes struggle to capture sufficient light to see, must overcome a number of fundamental physical limitations in order to extract reliable information from what are inherently dim and very noisy visual images. Recent research, particularly on the tropical sweat bee *Megalopta genalis* (a nocturnal bee native to the rainforests of Central America; Fig. 1), has revealed that nocturnal insects employ several neural strategies for this purpose. It is these strategies that have provided the inspiration for a new night-vision algorithm that dramatically improves the quality of video recordings made in very dim light [46], [47]. However, before explaining the algorithm and its performance, we will begin by describing the fundamental challenges faced by visual systems, and man-made imaging systems, in very dim light.

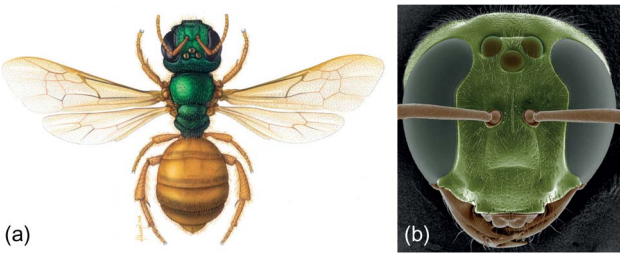

Fig. 1. *(a) The central American sweat bee Megalopta genalis (Halictidae), whose sensitive apposition eyes allows it to forage at night by visually learning landmarks along the foraging route and around the nest entrance. (b) A colored scanning electron microscope image of the head of Megalopta, showing the prominent compound eyes and the three ocelli.*

II. WHY IS VISION DIFFICULT IN DIM LIGHT?

From the outset it is important to point out that the nocturnal visual world is essentially identical to the diurnal visual world. The contrasts and colors of objects are almost identical. The only real distinguishing difference is the mean level of light intensity, which can be up to 11 orders of magnitude dimmer at night [48], [79], depending on the presence or absence of moonlight and clouds, and whether the habitat is open or closed (i.e., beneath the canopy of a forest). It is this difference that severely limits the ability of a visual system to distinguish the colors and contrasts of the nocturnal world. Indeed, many animals, especially diurnal animals, distinguish very little at all. In the end, the greatest challenge for an eye that views a dimly illuminated object is to absorb sufficient photons of light to reliably discriminate it from other objects [36], [77]–[79].

The main role of all photoreceptors is to transduce incoming light into an electrical signal whose amplitude is proportional to the light's intensity. The mechanism of transduction involves a complex chain of biochemical events within the photoreceptor that uses the energy of light to open ion channels in the photoreceptor membrane, thereby allowing the passage of charged ions and the generation of an electrical response. Photoreceptors are even able to respond to single photons of light with small but distinct electrical responses known as bumps [as they are called in the invertebrate literature; Fig. 2(a) and (b)]. At higher intensities, the bump responses fuse to create a graded response whose duration and amplitude is proportional to the duration and amplitude of the light stimulus. At very low-light levels, a light stimulus of constant intensity will be coded as a train of bumps generated in the retina at a particular rate, and at somewhat higher light levels, the constant intensity will be coded by a graded

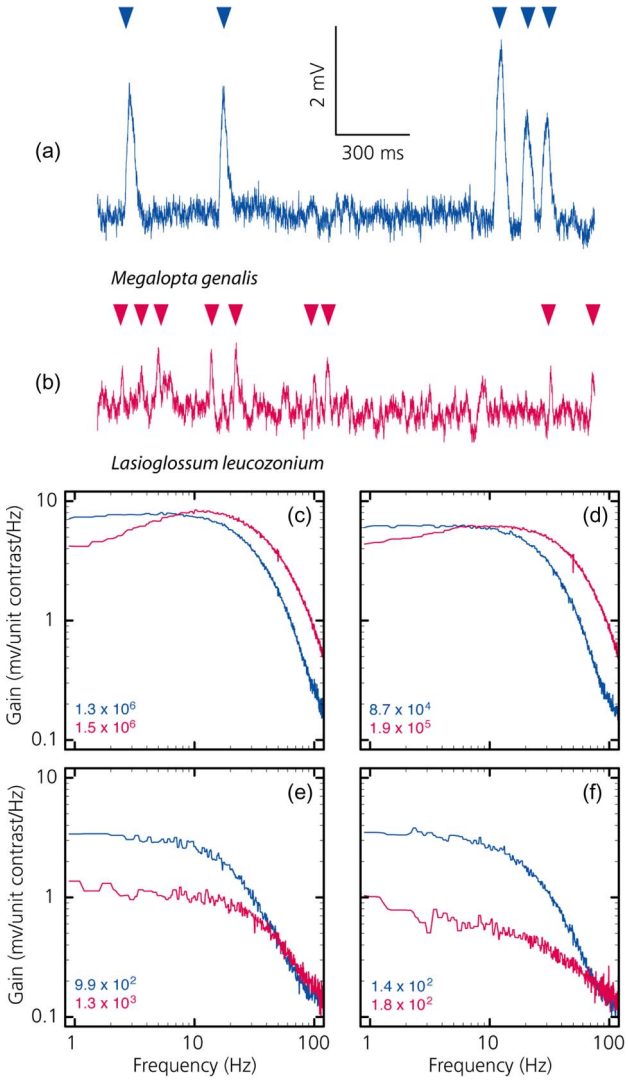

Fig. 2. ***Adaptations for nocturnal vision in the photoreceptors of the nocturnal sweat bee Megalopta genalis, as compared to photoreceptors in the closely related diurnal sweat bee Lasioglossum leucozonium. (a)–(b) Responses to single photons (or photon bumps: arrowheads) recorded from photoreceptors in Megalopta (a) and Lasioglossum (b). Note that the bump amplitude is larger, and the bump time course much slower, in Megalopta than in Lasioglossum. (c)–(f) Average contrast gain as a function of temporal frequency in Megalopta (blue curves, $n = 8$ cells) and Lasioglossum (red curves, $n = 8$ cells) at different adapting intensities, indicated as effective photons per second in each panel for each species (for each species, each stimulus intensity was calibrated in terms of effective photons, that is, the number of photon bumps per second that the light source elicited, thereby eliminating the effects of differences in the light gathering capacity of the optics between the two species, which is about 27× [19]). In light-adapted conditions [(c) and (d)], both species reach the same maximum contrast gain per unit bandwidth, although Lasioglossum has broader bandwidth and a higher corner frequency (the frequency at which the gain has fallen off to 50% of its maximum). In dark-adapted conditions [(e) and (f)], Megalopta has a much higher contrast gain per unit bandwidth. All panels adapted with permission from [17].***

potential of particular amplitude. At the level of the photoreceptors, the reliability of vision is determined by the repeatability of this response: for repeated presentations of the same stimulus, the reliability of vision is maximal if the rate of bump generation, or the amplitude of the graded response, remains exactly the same for each presentation. In practice, this is never the case, especially in dim light. Why is this so? The basic answer is that the visual response (and, as a consequence, its repeatability) is degraded by visual noise. Part of this noise arises from the stochastic nature of photon arrival and absorption (governed by Poisson statistics): each sample of N absorbed photons (or signal) has a certain degree of uncertainty (or noise) associated with it ($\pm\sqrt{N}$ photons). The relative magnitude of this uncertainty (i.e., $N/\sqrt{N} = \sqrt{N}$) is greater at lower rates of photon absorption, and these quantum fluctuations set an upper limit to the visual signal-to-noise ratio (SNR) [34], [61], [75]. As light levels fall, the fewer the number of photons that are absorbed, the greater is the noise relative to the signal and the less that can be seen. This photon "shot noise" limits detection reliability and is equally problematic for artificial imaging systems, such as cameras, as it is for eyes.

There are also two other sources of noise that further degrade visual discrimination by photoreceptors in dim light. The first of these, referred to as transducer noise, arises because photoreceptors are incapable of producing an identical electrical response, of fixed amplitude, latency, and duration, to each (identical) photon of absorbed light. This source of noise, originating in the biochemical processes leading to signal amplification, degrades the reliability of vision, particularly at slightly brighter light levels when photon shot noise no longer dominates [39], [43], [44]. The second source of noise, referred to as dark noise, arises because of the occasional thermal activation of the biochemical pathways responsible for transduction, which even occur in perfect darkness [5]. These activations produce dark events and electrical responses that are indistinguishable from those produced by real photons, and these are more frequent at higher retinal temperatures. At very low-light levels, this dark noise can significantly contaminate visual signals. In insects and crustaceans, dark events seem to be rare, only around 10 every hour at 25 °C, at least in those species where it has been investigated [13], [15], [44]. But in nocturnal toad rods, the rate is much higher, 360 per hour at 20 °C [6], and this sets the ultimate limit to visual sensitivity [1], [2]. This thermal noise is also a severe limitation to signal reliability in digital imaging systems.

How are the limitations imposed by noise overcome in visual systems? It turns out that several adaptations, both optical and neural, have evolved for improving visual reliability in dim light. These adaptations are surprisingly elegant and simple, and each contributes to one of three main mechanisms for improving visual reliability: 1) the possession of eyes with highly sensitive optics for

collecting as much of the available light as possible; 2) an enhancement of the neural image created in the retina; and 3) a subsequent optimal filtering of this image in space and time at a higher level in the visual system.

The first of these mechanisms involves the possession of an eye type that maximizes light capture. In nocturnal insects, including most moths and many beetles, this eye type is typically a refracting superposition compound eye. Like all compound eyes, superposition eyes (one of the two main types of compound eye; see Fig. 3) are composed of individual optical units known as ommatidia [Fig. 3(a)]. These units are tubular in shape and consist of one or more lenses: typically an outer transparent cuticular corneal facet lens (which is typically hexagonal in shape, as seen from the outside of the eye) and an inner crystalline cone. These supply light to a number of photoreceptor cells (or retinular cells) assembled underneath, each of which contributes to a cylindrical photoreceptive structure known as the rhabdom. Each ommatidium receives light from a small region of space—for most compound eyes somewhere between 1° and 10° across—and two neighboring ommatidia receive light from two neighboring such regions. Thus, the greater the number and density of ommatidia in a compound eye, the more finely sampled is visual space. In superposition eyes, single photoreceptors in the retina receive focused light from hundreds (and, in some extreme cases, thousands) of corneal facet lenses [Fig. 3(b)]. This eye type represents a vast improvement in sensitivity over the apposition compound eye [Fig. 3(c)], an eye type in which single photoreceptors receive light only from the single corneal facet lens residing in the same ommatidium. Not surprisingly, apposition eyes are typical of diurnal insects active in bright sunlight, and this includes all diurnal bees. Strangely, apposition eyes are also found in nocturnal bees such as *Megalopta*. Even stranger, despite the poor sensitivity afforded by apposition eyes, nocturnal bees see extremely well, with documented abilities to distinguish color [66] and to learn visual landmarks during foraging and homing [77]. How then do these nocturnal insects manage these remarkable feats? As we will now go on to see, it is the second and third mechanisms for improving visual reliability in dim light that are likely to be responsible: enhancement of the neural image created in the retina and spatio–temporal filtering of this image at a higher level in the visual system. It is these mechanisms we have now employed in the development of a new night-vision algorithm that dramatically improve the quality of video recordings made in very dim light, the details of which are described in Section IV.

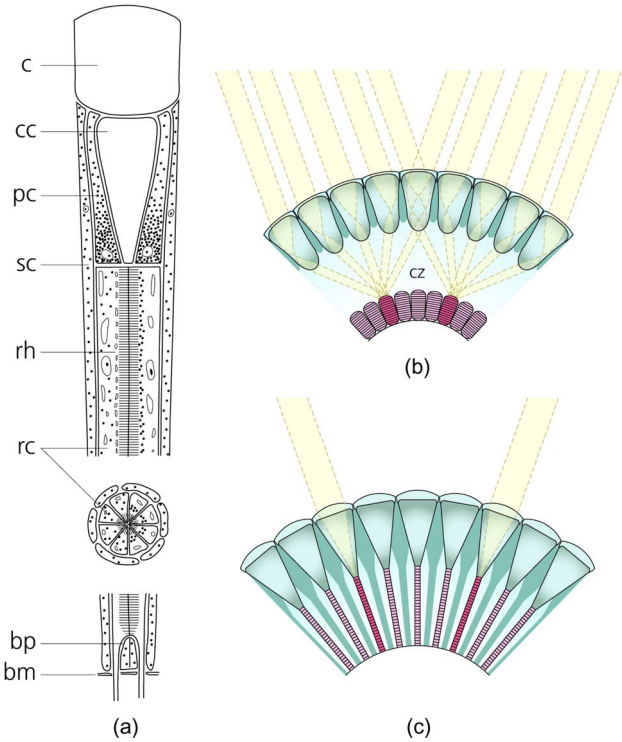

Fig. 3. *Compound eyes. (a) A schematic longitudinal section (and an inset of a transverse section) through a generalized Hymenopteran ommatidium, showing the corneal lens (c), the crystalline cone (cc), the primary pigment cells (pc), the secondary pigment cells (sc), the rhabdom (rh), the retinular cells (rc), the basal pigment cells (bp), and the basement membrane (bm). The left half of the ommatidium shows screening pigment granules in the dark-adapted state, while the right half shows them in the light-adapted state. Each ommatidium is responsible for reading the average intensity, color, and (in some cases) plane of polarization within the small region of space that it views. Two neighboring ommatidia view two neighboring regions of space. Thus, each ommatidium supplies a pixel of information to a larger image of pixels that the entire compound eye constructs. Larger compound eyes with more ommatidia thus have the potential for greater spatial resolution. Redrawn from [68]. (b) A refracting superposition compound eye (showing nine ommatidia in cross section). A large number of corneal facets and bullet-shaped crystalline cones collect and focus light—across the clear zone of the eye (cz)—toward single photoreceptors in the retina. Several hundred, or even thousands, of facets service a single photoreceptor. Not surprisingly, many nocturnal and deep-sea animals have refracting superposition eyes, and benefit from the significant improvement in sensitivity. Diagrams courtesy of D.-E. Nilsson. (c) A focal apposition compound eye (showing nine ommatidia in cross section). Light reaches the photoreceptors exclusively from the small corneal lens located directly above. This eye type is typical of day-active insects. Adapted from [77].*

III. NOCTURNAL VISION IN THE TROPICAL SWEAT BEE *MEGALOPTA GENALIS*

The nocturnal bee *Megalopta genalis* (Fig. 1) is a semi-social bee of the family Halictidae (the sweat bees) that lives in groups of up to 11 females in hollowed-out sticks suspended in the rainforest understory [83], [84]. A nocturnal lifestyle is rare in bees, and has occurred during evolution within only a few extant lineages [72], [80], and mostly in the tropical regions of the world. Reduced competition for floral nectar resources and the reduced risk of predation are

thought to be the main reasons why nocturnality in bees has arisen [83]. *Megalopta* leaves its nest and forages on flowers under the thick rainforest canopy during two short time windows after dusk and before dawn, when light levels are comparable to starlight levels above the forest canopy [32], [77]. Vision plays a central role in this behavior, both in the control of flight and in the learning of spatial landmarks used during homing [3], [77].

Despite the difficulties imposed by its nocturnal lifestyle, the visual system of *Megalopta* is remarkably well adapted for vision at night. Optically, their eyes are around 27 times more sensitive to light than the strictly day-active European honeybee *Apis mellifera*, an improvement that results from having facet lenses almost twice as wide as in *Apis*, and considerably larger rhabdoms having 16 times the photon capture area. These morphological differences allow each ommatidium in *Megalopta* to collect light from regions of space that are around seven times wider than in *Apis* [77].

Even though an optical sensitivity improvement of 27 times sounds impressive, compared to the 100 million times difference in light intensity experienced by *Apis* and *Megalopta*, the improvement is modest at best, and highly unlikely on its own to bridge the massive gap required for *Megalopta* to see well in very dim light. The only remaining option is to dramatically improve visual reliability—beyond the limitations imposed by the optics—by employing neural mechanisms that optimally process the incoming visual information. In insects, these mechanisms have been found at the very first stages of light detection in the photoreceptors of the retina, as well as at higher levels of visual processing in the brain. As it turns out, the neural principles that have been discovered are directly applicable to the enhancement of video sequences collected in very dim light. Indeed, as we will see in Section IV, they are essential elements of our new night-vision algorithm.

A. Principles of Biological Photon Detection in Nocturnal Insects

As we mentioned before, most animal photoreceptors are capable of detecting single photons of light, responding to them with small but discrete voltage responses that in insects are known as bumps [Fig. 2(a) and (b)]. Research has revealed that these bumps are much larger in nocturnal and crepuscular insects than in diurnal insects [17], [40], revealing an intrinsic benefit for vision in dim light. From electrophysiological recordings of photoreceptors in two closely related nocturnal and diurnal halictid bee species—the nocturnal *Megalopta genalis* [Fig. 2(a)] and the diurnal *Lasioglossum leucozonium* [Fig. 2(b)]—bump amplitude in the nocturnal species was found to be much greater than in the diurnal species [17]. These larger bumps have also been reported from other nocturnal and crepuscular arthropods (e.g., crane flies, cockroaches, and spiders [28], [38], [40], [55]), and in locusts, which are considered to be diurnal insects, bump size can even vary at different times of the day, becoming significantly larger at night [26]. Larger bumps indicate that the photoreceptor's gain of transduction is greater, and thus in *Megalopta* the gain of transduction is greater than in *Lasioglossum*. This higher transduction gain manifests itself as a higher contrast gain, which means that the photoreceptor has a greater voltage response per unit change in light intensity (or contrast). At lower light levels the contrast gain is up to five times higher in *Megalopta* than in *Lasioglossum*, which results in greater signal amplification and the potential for improved visual reliability in dim light [Fig. 2(c)–(f)]. One problem with having a higher gain is that it not only elevates the visual signal, but also it elevates several sources of visual noise by the same amount. Thus, for this strategy to pay off, subsequent stages of processing are needed to reduce the noise. As we will see in Section III-B, this processing involves a summation of photoreceptor responses in space and time.

In addition to a higher contrast gain, the photoreceptors of nocturnal insects tend also to be significantly slower than those of their day-active relatives. Despite compromising temporal resolution, slower vision in dim light (analogous to having a longer exposure time on a camera) is beneficial because it increases the visual SNR and improves contrast discrimination at lower temporal frequencies by suppressing photon noise at frequencies that are too high to be reliably resolved [22], [23]. Temporal resolution can be measured in several ways, for instance, by measuring the ability of a photoreceptor to follow a light source whose intensity modulates sinusoidally over time: a photoreceptor that can follow a light source modulating at a high frequency is considered to be fast. Thus, in the frequency domain, slower vision is equivalent to saying that the temporal bandwidth is narrow, or more precisely, that the temporal corner frequency is low. This is often defined as the frequency where the response has fallen to 50% of its maximum value (i.e., by 3 dB), and lower values indicate slower vision. In *Megalopta*, it is around 7 Hz in dark-adapted conditions. In diurnal *Lasioglossum*, the dark-adapted corner frequency is nearly three times the value found in *Megalopta*, around 20 Hz, a value that is nonetheless considerably less than that typical of the diurnal, highly maneuverable, and rapidly flying higher flies (50–107 Hz; [40]). The difference in temporal properties between the two bee species is most likely due to different photoreceptor sizes, and different numbers and types of ion channels in the photoreceptor membrane [37], [40], [50], [51], [85].

B. Spatial and Temporal Summation

As mentioned above, even though slow photoreceptors of high contrast gain are potentially beneficial for improving visual reliability in dim light, these benefits may not be realized because of the contamination of visual signals by various sources of visual noise. The solution to overcome the problems of noise contamination is to neurally sum visual signals in space and time [35], [36], [63]–[65], [76].

We have already discussed summation of photons in time before: when light gets dim, the slower photoreceptors of nocturnal animals can improve visual reliability by integrating signals over longer periods of time [22], [23], [36]. Even slower vision could be obtained by neurally integrating (summing) signals at a higher level in the visual system. However, temporal summation has a cost: the perception of fast-moving objects is seriously degraded. This is potentially disastrous for a fast-flying nocturnal animal (like a nocturnal bee or moth) that needs to negotiate obstacles. Not surprisingly, temporal summation is more likely to be employed by slowly moving animals.

Summation of visual signals in space can also improve image quality. Instead of each visual channel collecting photons in isolation (as in bright light), the transition to dim light could activate specialized laterally spreading neurons which couple the channels together into groups. Each summed group—themselves now defining the channels—could collect considerably more photons over a much wider visual angle, albeit with a simultaneous and unavoidable loss of spatial resolution. Despite being much brighter, the image would become necessarily coarser. In *Megalopta*, such laterally spreading neurons have been found in the first optic ganglion (lamina ganglionaris), the first visual processing station in the brain. The bee's four classes of lamina monopolar cells (or L-fibers, L1–L4), which are responsible for the analysis of photoreceptor signals arriving from the retina, are housed within each neural cartridge of the lamina, a narrow cylinder of lamina tissue that resides below each ommatidium. Compared to the L-fibers of the diurnal honeybee *Apis mellifera*, those of *Megalopta* have lateral processes that extensively spread into neighboring cartridges [Fig. 4(a)]: cells L2, L3, and L4 spread to 12, 11, and 17 lamina cartridges, respectively, while the homologous cells in *Apis* spread, respectively, to only two, zero, and four cartridges [20], [21]. Similar laterally spreading neurons have also been found in other nocturnal insects, including cockroaches [59], fireflies [52], and hawkmoths [69]. Spatial and temporal summation strategies have the potential to greatly improve the visual SNR in dim light (and thereby the reliability of vision) for a narrower range of spatial and temporal frequencies [33], [71], [76]. However, despite the consequent loss in spatial and temporal resolution, summation would tend to average out the noise (which is uncorrelated between channels) or significantly reduce its amplitude. Thus, summation would maximize nocturnal visual reliability for the slower and coarser features of the world. Those features that are faster and finer, and inherently noisy, would be filtered out. However, it is, of course, far better to see a slower and coarser world than nothing at all.

These conclusions can be visualized theoretically [76]. Both *Megalopta* [Fig. 4(b)] and *Apis* [Fig. 4(c)] are able to resolve spatial details in a scene at much lower intensities with summation than without it [71]. These theoretical results assume that both bees experience an angular velocity of 240°/s during flight, a value that has been measured from high-speed films of *Megalopta* flying at night. At the lower light levels where *Megalopta* is active, the optimum visual performance shown in Fig. 4(b) is achieved with an integration time of about 30 ms and summation from about 12 ommatidia (or cartridges). This integration time is close to the photoreceptor's dark-adapted value [77], and the extent of predicted spatial summation is very similar to the number of cartridges to which the L2 and L3 cells actually

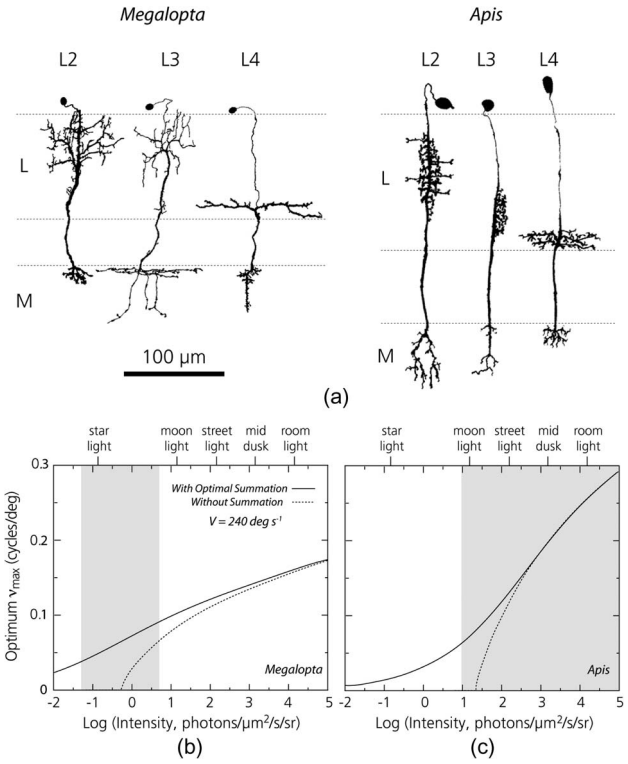

Fig. 4. ***Spatial summation in nocturnal bees. (a) Comparison of the first-order interneurons, L-fiber types L2, L3, and L4, of the Megalopta genalis female (left) and the worker honeybee Apis mellifera (right). Compared to the worker honeybee, the horizontal branches of L-fibers in the nocturnal halictid bee connect to a much larger number of lamina cartridges, suggesting a possible role in spatial summation. L = lamina, M = medulla. Reconstructions from Golgi-stained frontal sections. Adapted from [21] and [60]. (b)–(c) Spatial and temporal summation modeled at different light intensities in Megalopta genalis (b) and Apis mellifera (c) for an image velocity of 240°/s (measured from Megalopta genalis during a nocturnal foraging flight [77]). Light intensities are given for 540 nm, the peak in the bee's spectral sensitivity. Equivalent natural intensities are also shown. The finest spatial detail visible to flying bees (as measured by the maximum detectable spatial frequency ν_{max}) is plotted as a function of light intensity. When bees sum photons optimally in space and time (solid lines), vision is extended to much lower light intensities (nonzero ν_{max}) compared to when summation is absent (dashed lines). Note that nocturnal bees can see better in dimmer light than honeybees. Gray areas denote the light intensity window within which each species is normally active (although honeybees are also active at intensities higher than those presented on the graph). Figure reproduced from [79] with permission from Elsevier.***

branch [21], thus strengthening the hypothesis that the lamina monopolar cells are involved in spatial summation. Interestingly, even in the honeybee *Apis*, summation can improve vision in dim light [Fig. 4(c)]. The Africanised subspecies *Apis mellifera scutellata* and the closely related southeast Asian giant honeybee *Apis dorsata* both forage on large pale flowers during dusk and dawn, and even throughout the night, if a half-full or larger moon is present in the sky. This ability can be explained only if bees optimally sum photons over space and time [82], and this is also revealed in Fig. 4(c) (for an angular velocity of 240°/s). At the lower light levels where *Apis* is active, the optimum visual performance shown in Fig. 4(c) is achieved with an integration time of about 18 ms and summation from about three or four cartridges. As in *Megalopta*, this integration time is close to the photoreceptor's dark-adapted value, and the extent of predicted spatial summation is again very similar to the number of cartridges to which the L2 and L3 cells actually branch.

The limitations in detection reliability experienced by nocturnal eyes, and the biological means they use to overcome them, lie at the heart of our new night-vision algorithm. As we will now go on to show, this algorithm radically improves the quality of video sequences collected in dim light. The remainder of this review summarizes both the algorithm and the results.

IV. A BIOLOGICALLY INSPIRED NIGHT-VISION ALGORITHM

In this section, we will describe our video enhancement algorithm, and relate it to the biological model that it is based on. We will give an overview of our approach, and then, in Sections IV-C–IV-G, the different parts of the algorithm will be described in detail. But we will start by describing some of the related work that has been done in the field of image denoising and low-light level vision.

A. Related Work

There exists a multitude of noise reduction techniques that apply spatio–temporal weighted averaging for noise reduction purposes. Many authors have additionally realized the benefit of trying to reduce the noise by filtering the sequences along motion trajectories in spatio–temporal space, and, in this way, ideally avoid motion blur and unnecessary amounts of spatial averaging. These noise reduction techniques are usually referred to as motion-compensated (spatio–)temporal filtering. In [29], means along the motion trajectories were calculated, while in [53] and [49], weighted averages, dependent on the intensity structure and noise in a small neighborhood, were applied. In [62], so-called linear minimum mean square error filtering was used along the trajectories. Most of the motion-compensated methods use some kind of block-matching technique for the motion estimation that usually, for efficiency, employs rather few images, commonly two or three. However, in [62], [53], and [16], variants using optical flow estimators were presented. A drawback of the motion-compensating methods is that the filtering relies on a good estimate of motion to give a good output, without excessive blurring. The motion estimation is especially complicated for sequences severely degraded by noise. Different approaches have been applied to deal with this problem, often by simply reducing the influence of the temporal filter in difficult areas. This, however, often leads to disturbing noise at, for example, object contours.

An interesting family of smoothing techniques for noise reduction are those that solve an edge-preserving anisotropic diffusion equation on the images, i.e., modeling the noise reduction problem using a version of the heat equation. This approach was pioneered by Perona and Malik [54] and has had many successors, including the work by Weickert [86]. These techniques have also been extended to 3-D and spatio–temporal noise reduction in video; cf., [74] and [42]. In [74] and [87], the so-called structure tensor or second moment matrix is applied in a similar manner to our approach in order to analyze the local spatio–temporal intensity structure and steer the smoothing accordingly. The drawbacks of techniques based on diffusion equations include the fact that the solution has to be found using an often time-consuming iterative procedure. Moreover, it is very difficult to find a suitable stopping time for the diffusion, at least in a general and automatic manner. These drawbacks make these approaches unsuitable for video processing.

A better approach is to apply single-step structure-sensitive adaptive smoothing kernels. The bilateral filters introduced by Tomasi and Manduchi [73] for 2-D images fall within this class. Here, edges are maintained by calculating a weighted average at every point using a Gaussian kernel, where the coefficients in the kernel are attenuated based on how different the intensities are in the corresponding pixels compared to the center pixel. This makes the local smoothing very dependent on the correctness in intensity in the center pixel, which cannot be assumed in images heavily disturbed by noise.

An approach that is closely connected to both bilateral filtering and to anisotropic diffusion techniques based on the structure tensor is the structure-adaptive anisotropic filtering presented by Yang *et al.* [88]. We base our spatio–temporal smoothing on the ideas in this paper, but extend it to our special setting for video and very low-light levels. For a study of the connection between anisotropic diffusion, adaptive smoothing, and bilateral filtering, see [4].

Some of the most successful approaches to image denoising over the recent years belong to the group of block-matching algorithms. These include the nonlocal means algorithms [8] and the popular Block Matching 3D (BM3D) approach [9], [10]. BM3D is the algorithm that we have tried which has the most comparable result to our own.

Another group of algorithms connected to our work are in the field of high dynamic range (HDR) imaging;

cf., [30], [58], and the references therein. The aim of HDR imaging is to alleviate the restriction caused by the low dynamic range of ordinary charge-coupled device (CCD) cameras, i.e., the restriction to the ordinary 256 intensity levels for each color channel. Most HDR imaging algorithms are based on using multiple exposures of a scene with different settings for each exposure and then using different approaches for storing and displaying these extended image data. However, the HDR techniques are not especially aimed at the kind of low-light level data that we are targeting in this work, where the utilized dynamic range in the input data is in the order of 5 to 20 intensity levels and the SNR is extremely low.

Surprisingly few published studies exist that especially target noise reduction in low-light-level video. Two examples are the aforementioned method by Bennett and McMillan [7] and the technique presented by Lee *et al.* [41]. In [41], very simple operations are combined in a system presumably developed for easy hardware implementation in, for example, mobile phone cameras and other compact digital video cameras, and this method cannot handle the high levels of noise that we target in this work.

The approach taken by Bennett and McMillan [7] for low dynamic range image sequences is more closely connected to our technique. Their virtual exposure framework includes the bilateral adaptive spatio–temporal accumulation (ASTA) filter and a tone mapping technique. The ASTA filter, which changes to relatively more spatial filtering in favor of temporal filtering when motion is detected, is in this way related to the biological model described in Section III. However, since bilateral filters are applied, the filtering is edge sensitive and the temporal bilateral filter is additionally used for the local motion detection. The filters apply novel dissimilarity measures to deal with the noise sensitivity of the original bilateral filter formulation. The filter size applied at each point is decided by an initial calculation of a suitable amount of amplification using tone mapping of an isotropically smoothed version of the image. A drawback of the ASTA filter is that it requires global motion detection as a preprocessing step to be able to deal with moving camera sequences. The sequence is then warped to simulate a static camera sequence.

B. Overview of Algorithm

We will now give an overview of the basic steps in our algorithm. We will try to follow the methodology described in Section III to optimally enhance our image sequence. We start with a low-light video sequence $I_{in}(\mathbf{x})$, where $\mathbf{x} \in \mathbb{R}^3$ is a spatio–temporal point. One frame of an example is given on the left-hand side of Fig. 5. The input gray values (or color values) are amplified using a nonlinear amplification function T so that $I_T(\mathbf{x}) = T(I_{in}(\mathbf{x}))$. Function T could be fixed or adapted to the input values. Some examples are discussed in Section IV-C. Since we not only amplify the signal but also the noise, we get, in general, a very noisy output video sequence. This relates to the amplification described in Section III-A. One frame of an amplified sequence is given in the middle of Fig. 5. Now comes the difficult part, namely to reduce the noise but keep the signal intact. In order to do this, we must in some way distinguish the noise from the signal. We will try to mimic the ideas in Section III-B where the signal is retrieved through optimally summing in space and time. Almost all noise reduction algorithms are based on averaging the signal. But we do not want to average pixels or voxels that have different signal value, i.e., we do not want to average over edges in the signal structure. Since we are dealing with a video sequence, edges appear for two different reasons: 1) due to physical appearance differences in objects, such as boundaries of objects or texture and color changes; and 2) due to motion of either the camera or objects.

Fig. 5. *(Left) Frame from a very dark input sequence. (Middle) One frame of the amplified sequence. Here, we have structures clearly visible, but the noise in the image has also been amplified. (Right) One frame from the final output of Algorithm 1. The structures are preserved but the noise has been highly suppressed.*

We will consider the image sequence as a spatio–temporal volume, as depicted in the top part of Fig. 6. Our goal is to construct smoothing kernels that adapt to the local spatio–temporal intensity structure. These kernels should be large in directions where there are no edges, and small in directions where we have strong edges. In the bottom part of Fig. 6, the kernels for two different points are shown. The man walking gives rise to an intensity streak in the image stack. The estimated kernels adapt to this local spatio–temporal structure so that we do not get smoothing over edges. Note that the kernel at the point with more structurally isotropic neighborhood is quite isotropic, whereas the kernel at the point with the man walking is elongated in the direction of the intensity streak.

We will use the so-called structure tensor to estimate the local spatio–temporal structure at each point and use this to construct the smoothing kernels. This will also include calculating the eigenvalues and eigenvectors that we use to orient and stretch our kernels. The details are given in Section IV-D. When we have constructed the kernels we can estimate the new gray values by integration (or summation as we are dealing with discrete structures). The different steps are summarized in Algorithm 1.

Algorithm 1: Low-Light Video Enhancement

1: Given a low intensity input video sequence

$$I_{\text{in}}(\mathbf{x}), \qquad \mathbf{x} \in D \subset \mathbb{R}^3.$$

2: Apply an amplifying intensity transformation $T : \mathbb{R} \to \mathbb{R}$

$$I_T(\mathbf{x}) = T(I_{\text{in}}(\mathbf{x})).$$

3: **for** each $\mathbf{x_0} \in D$ and a neighborhood Ω **do**
4: Calculate the structure tensor $\mathbf{J}(\mathbf{x_0})$.
5: Calculate the eigenvalues and eigenvectors of $\mathbf{J}(\mathbf{x_0})$.
6: Construct the summation kernel $k : \mathbb{R}^2 \to \mathbb{R}$

$$k(\mathbf{x_0}, \mathbf{x}), \qquad \mathbf{x} \in \Omega.$$

7: Integrate the output intensity

$$I_{\text{out}}(\mathbf{x_0}) = \iiint_{\Omega} k(\mathbf{x_0}, \mathbf{x}) I_T(\mathbf{x}) d\mathbf{x}.$$

8: **end for**

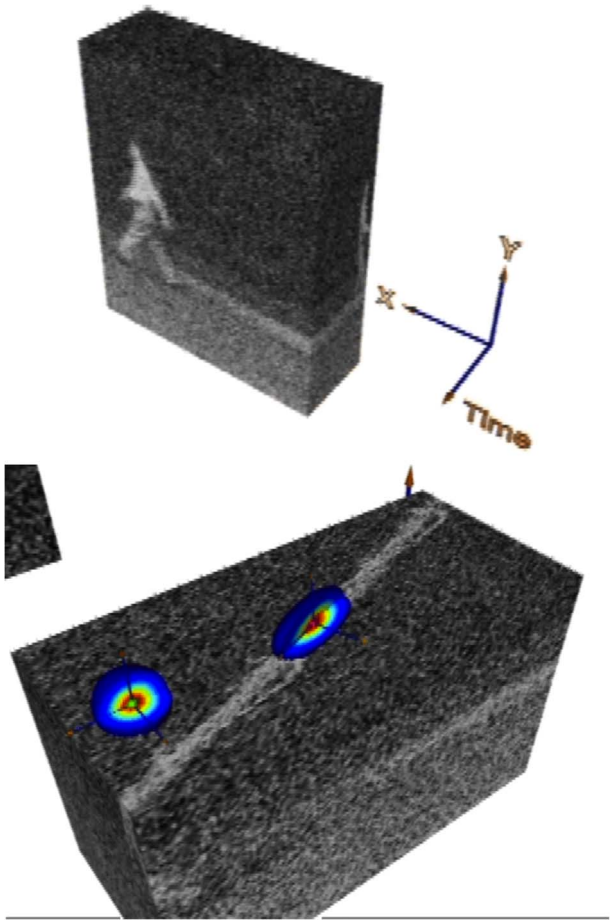

Fig. 6. *The image stack and the smoothing kernel construction for the image sequence in Fig. 5. In the top part, we see the image stack with the corresponding coordinate system. In the bottom part, we have cut the stack at a fixed y-coordinate and superimposed the kernels at two different points. The man walking gives rise to an intensity streak in the image stack. The estimated kernels adapt to this local spatio–temporal structure so that we do not get smoothing over edges. Note that the kernel at the point with the more structurally isotropic neighborhood is quite isotropic, whereas the kernel at the point with the man walking is elongated in the direction of the intensity streak.*

C. Intensity Transformation

The procedure of intensity transformation is also commonly referred to as tone mapping. The tone mapping could actually be performed either before or after the noise reduction, with a similar output, as long as the parameters for the smoothing kernels are chosen to fit the current noise level.

In the virtual exposures method of Bennett and McMillan [7], a tone mapping procedure is applied where the actual mapping is a logarithmic function similar to the one proposed by Drago *et al.* [14]. The tone mapping procedure also contains additional smoothing using spatial and temporal bilateral filters and an attenuation of details, found by the subtraction of a filtered image from a nonfiltered image. We instead choose to do all smoothing in the separate noise reduction stage and here concentrate on the tone mapping.

The tone mapping procedure of Bennett involves several parameters, both for the bilateral smoothing filters and for changing the acuteness of two different mapping functions, one for the large scale data and one for the attenuation of the details. These parameters have to be set manually and will not adapt if the lighting conditions change in the image sequence. Since we aim for an automatic procedure, we instead opt for a modified version of the well-known procedure of histogram equalization; cf., [18]. Histogram equalization is parameter free and increases the contrast in an image by finding a tone mapping that evens out the intensity histogram of the input image as much as possible. However, for many images, histogram equalization gives a too extreme mapping, which, for example, saturates the brightest intensities so that structure information here is lost. It also heavily changes the local intensity distribution in ways that change the appearance of the images. Therefore, we apply contrast-limited histogram equalization as presented by Pizer *et al.* [56], but without the tiling that applies different mappings to different areas (tiles) in the image. In the contrast-limited histogram equalization, a parameter, the clip-limit β,

sets a limit on the derivative of the slope of the mapping function. If the mapping function, found by histogram equalization, exceeds this limit, the increase in the critical areas is spread equally over the mapping function.

D. Spatio–Temporal Filtering

We will now present our modifications and extensions to the structure-adaptive anisotropic image filtering by Yang *et al.* [88] in order to improve its applicability and make it suitable for our low-light-level vision objective.

A new image $I_{out}(\mathbf{x_0})$ is obtained by applying at each spatio–temporal point $\mathbf{x_0} = (x_0, y_0, t_0)$ a kernel $k(\mathbf{x_0}, \mathbf{x})$ to the original image $I_{in}(\mathbf{x_0})$ such that

$$I_{out}(\mathbf{x_0}) = \frac{1}{\mu(\mathbf{x_0})} \iiint_{\Omega} k(\mathbf{x_0}, \mathbf{x}) I_{in}(\mathbf{x}) d\mathbf{x} \tag{1}$$

where

$$\mu(\mathbf{x_0}) = \iiint_{\Omega} k(\mathbf{x_0}, \mathbf{x}) d\mathbf{x} \tag{2}$$

is a normalizing factor. The normalization makes the sum of the kernel elements equal to 1 in all cases, so that the mean image intensity does not change. The area Ω over which the integration, or in the discrete case, summation, is made is chosen as a finite neighborhood centered around $\mathbf{x_0}$, typically 13^3 voxels.

Since we want to adapt the filtering to the spatio–temporal intensity structure at each point, in order to reduce blurring over spatial and temporal edges, we calculate a kernel $k(\mathbf{x_0}, \mathbf{x})$ individually for each point $\mathbf{x_0}$. The kernels should be wide in directions of homogeneous intensity and narrow in directions with important structural edges. To find these directions, the intensity structure is analyzed by the so-called structure tensor or second moment matrix. This object has been developed and applied in image analysis in numerous papers, e.g., by Jähne [27]. The tensor $\mathbf{J}_\rho(\mathbf{x_0})$ is defined in the following way:

$$\mathbf{J}_\rho(\nabla I(\mathbf{x_0})) = G_\rho \star \left(\nabla I(\mathbf{x_0}) \nabla I(\mathbf{x_0})^T\right) \tag{3}$$

where

$$\nabla I(\mathbf{x_0}) = \left[\frac{\partial I}{\partial x_0} \frac{\partial I}{\partial y_0} \frac{\partial I}{\partial t_0}\right]^T \tag{4}$$

is the spatio–temporal intensity gradient of I at point $\mathbf{x_0}$. G_ρ is the Gaussian kernel function

$$G_\rho(\mathbf{x}) = \frac{1}{\mu} e^{-\frac{1}{2}\left(\frac{x^2+y^2+t^2}{\rho^2}\right)} \tag{5}$$

where μ is the normalizing factor. The notation $\star$ means elementwise convolution of the matrix $\nabla I(\mathbf{x_0}) \nabla I(\mathbf{x_0})^T$ in a neighborhood centered at $\mathbf{x_0}$. It is this convolution that gives us the smoothing in the direction of gradients, which is the key to the noise insensitivity of this method.

Eigenvalue analysis of $\mathbf{J}_\rho$ will now give us the structural information that we seek. The eigenvector $\mathbf{v_1}$ corresponding to the smallest eigenvalue λ_1 will be approximately parallel to the direction of minimum intensity variation while the other two eigenvectors are orthogonal to this direction. The magnitude of each eigenvalue will be a measure of the amount of intensity variation in the direction of the corresponding eigenvector. For a deeper discussion on eigenvalue analysis of the structure tensor, see [24].

The basic form of the kernels $k(\mathbf{x_0}, \mathbf{x})$ that are constructed at each point $\mathbf{x_0}$ is that of a Gaussian function

$$k(\mathbf{x_0}, \mathbf{x}) = e^{-\frac{1}{2}(\mathbf{x}-\mathbf{x_0})^T R \Sigma^2 R^T (\mathbf{x}-\mathbf{x_0})} \tag{6}$$

including a rotation matrix R and a scaling matrix Σ. The rotation matrix is constructed from the eigenvectors $\mathbf{v_i}$ of $\mathbf{J}_\rho$

$$R = [\mathbf{v_1} \quad \mathbf{v_2} \quad \mathbf{v_3}] \tag{7}$$

while the scaling matrix has the following form:

$$\Sigma = \begin{bmatrix} \frac{1}{\sigma(\lambda_1)} & 0 & 0 \\ 0 & \frac{1}{\sigma(\lambda_2)} & 0 \\ 0 & 0 & \frac{1}{\sigma(\lambda_3)} \end{bmatrix}. \tag{8}$$

The function $\sigma(\lambda_i)$ is a decreasing function that sets the width of the kernel along each eigenvalue direction. The theory in [88] is mainly developed for 2-D images, and measures of corner strength and of anisotropism, both involving ratios of the maximum and minimum eigenvalues, are there calculated at every point x_0. An extension of this to the 3-D case is then discussed. However, we have not found these two measures to be adequate for the 3-D

case since they give too much focus on singular corner points in the video input and to a large extent disregard the linear and planar structures that we want to preserve in the spatio–temporal space. For example, a dependence of the kernel width in the temporal direction on the eigenvalues corresponding to the spatial directions does not seem appropriate in a static background area. We instead simply let an exponential function depend directly on the eigenvalue λ_i in the current eigenvector direction $\mathbf{v}_i$ in the following way:

$$\sigma(\lambda_i, \mathbf{x_0}) = \begin{cases} \Delta\sigma e^{-\frac{\lambda_i}{d}+\lambda_{\min}} + \sigma_{\min}, & \lambda_i > \lambda_{\min} \\ \sigma_{\max}, & \lambda_i \leq \lambda_{\min} \end{cases} \quad (9)$$

where $\Delta\sigma = (\sigma_{\max} - \sigma_{\min})$, so that σ_i attains its maximum $\sigma_{\max}$ below $\lambda = \lambda_{\min}$ and asymptotically approaches its minimum $\sigma_{\min}$ when $\lambda \to \infty$. Parameter d scales the width function along the λ-axis and has to be set in relation to the current noise level. Since the part of the noise that stems from the quantum nature of light, i.e., the photon shot noise, depends on the brightness level, it is signal dependent and parameter d should ideally be set locally. However, it has been noted that when the type of camera and the type of tone mapping are fixed, a fixed value on d usually works for a large part of the dynamic range. When changing the camera and tone mapping approach, a new value of d has to be found for optimal performance.

When the widths $\sigma(\lambda_i, \mathbf{x_0})$ have been calculated and the kernel subsequently constructed according to (6), (1) is used to calculate the output intensity I_{out} of the smoothing stage at the current pixel $\mathbf{x_0}$.

In biological vision, edges are often perceived stronger than they actually are. This effect, called lateral inhibition, is due to the ability of an excited neuron to reduce the response of its neighboring neurons. We have mimicked this behavior in an edge-sharpening version of our algorithm. In order to do this, we simply replace the Gaussian kernel with the following base kernel:

$$k_{hb}(\mathbf{x}) = \frac{1}{\mu} e^{-\frac{1}{2}\left(\frac{\mathbf{x}^T\mathbf{x}}{\sigma^2}\right)} (1 - \alpha \mathbf{x}^T \mathbf{x}) \quad (10)$$

which is stretched and rotated in the same way as before. In the top row of Fig. 7, the difference between the two kernels is shown. The bottom row shows the output of our algorithm using the two different kernels.

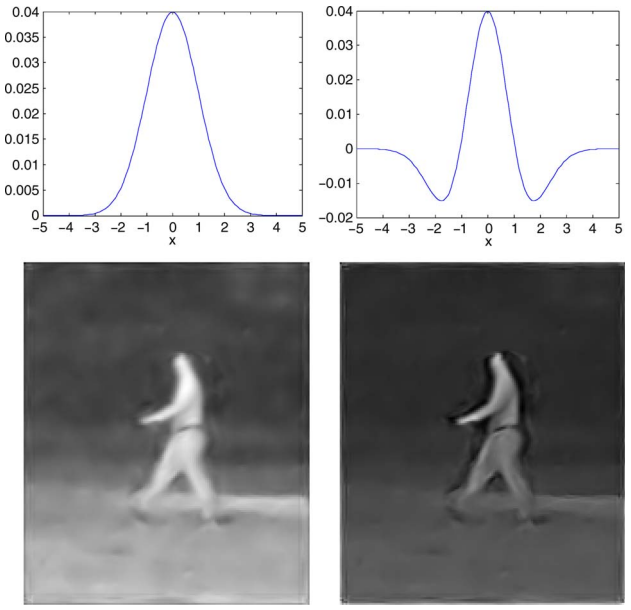

Fig. 7. *Impact of using the lateral inhibition technique described in Section III-B. The top row shows the basic kernels in one dimension used in the summation. To the left is a standard Gaussian and to the right sharpening version. The bottom row shows frames from the output of our algorithm. The left shows the result of running the algorithm with a Gaussian kernel, and the right shows the result using the sharpening kernel.*

E. Parallelization and Computational Aspects

The visual system of most animals is based on a highly parallel architecture. Looking at the overview of Algorithm 1 two things are clear. The algorithm is linear in the number of voxels in the input image sequence and it is highly parallelizable. The output of each voxel depends only on a neighborhood of that given voxel. The most expensive parts are the calculation of the structure tensor, including the gradient calculation, the elementwise smoothing, and the actual filtering, or summation. This is a task for which the graphics processing units of modern graphics cards are very well suited.

We have implemented the whole adaptive enhancement methodology as a combined central processing unit/graphics processing unit (CPU/GPU) algorithm. All image preprocessing and postprocessing is performed on the CPU. This includes image input/output and the amplification step. If we use histogram equalization as a contrast amplification, this requires summation over all pixels. This computation is not easily adapted to a GPU, as the summation would have to be done in multiple passes. However, as these steps constitute a small amount of the execution time, a simpler CPU implementation is adequate here.

In summary, by exploiting the massively parallel architecture of modern GPUs, we obtain interactive frame rates on a single nVidia GeForce 8800-series graphics card.

F. Adaptation of Parameters to Noise Levels

The amount of noise in an image sequence changes depending on the brightness level: the more SNR decreases, the darker the area gets. Since we want the algorithm to adapt to changing light levels, both spatially within the same image and temporally in the image sequence, the

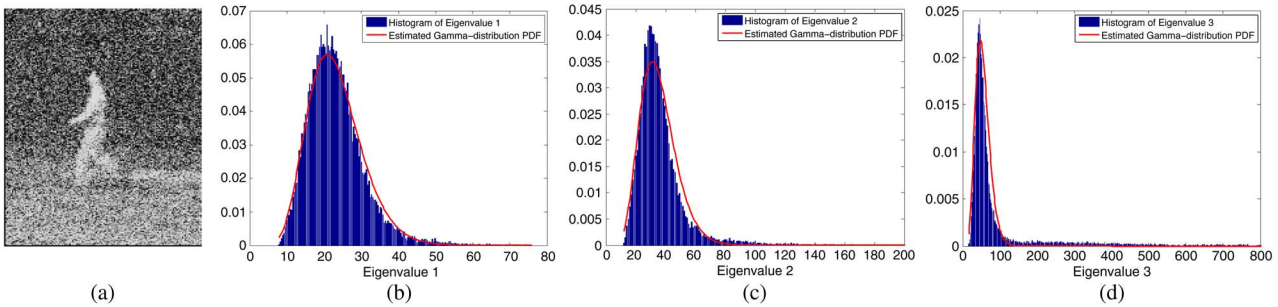

Fig. 8. ***Histogram of eigenvalues. (a) One frame of an amplified noisy image sequence. (b)–(d) Histograms of eigenvalues 1–3 in blue. Also shown in red are fitted gamma distributions to the data.***

width function needs to depend on the noise level and adapt to the SNR. This adaptation is governed by the measure d. In this section, we briefly discuss how the noise in the eigenvalues can be modeled, and this, in turn, gives us a way of choosing the parameters in the algorithm in an appropriate way.

Fig. 9. ***Illustration of results from four different input sequences. (a) One frame of the dark input sequence. (b) One frame of the highly noisy amplified image sequence. (c) One frame of our approach. (d) One frame of the result of running BM3D. The two top rows are taken by a standard compact digital camera, a Canon IXUS 800 IS. The third row is taken by a handheld consumer camcorder, a Sony DCR-HC23. The bottom row is taken by an AVT Marlin machine vision camera. The two top and the bottom sequences have large camera motions and the two bottom sequences have additional motion in the scene.***

If we model the image sequence as a signal with added independent identically distributed (i.i.d.) Gaussian noise of zero mean, we can say the following about the noise in the estimated eigenvalues. First, the sequence is filtered with Gaussian filters as well as differentiation filters to estimate the gradients in the sequence at every point. The noise in the estimated gradients will hence also be Gaussian but with a different variance. The structure tensor is calculated from the estimated gradients. As the structure tensor is made up of the outer product of the gradients, and we then filter the tensor elementwise, the tensor elements will be a sum of squares of Gaussian distributed variables. We will model the noise in the eigenvalues as gamma distributed. This is not the exact distribution since the calculation of the eigenvalues and the structure tensor includes a number of nonlinear steps, but we have seen in our experiments that we capture enough of the statistics using this approximation. In Fig. 8, histograms of the eigenvalues of the structure tensors for an image are shown in blue. We can fit a Gamma probability density function (pdf) to this histogram in a least squares manner. The resulting gamma distribution is shown in red. In our experience, the eigenvalues generally follow a gamma distribution well.

From the Gamma distribution, means $\bar{\lambda}_i$ and variances $s^2_{\lambda_i}$ can be estimated. These parameters can be used to set the parameters, i.e., d, of the algorithm.

G. Considerations for Color

The discussion so far has dealt with intensity images. We will here discuss some special aspects of the algorithm when it comes to processing color images.

In applying the algorithm to RGB color image data, one could envision a procedure where the color data in the images are first transformed to another color space including an intensity channel, e.g., the hue-saturation-value (HSV) color space; cf., [18]. The algorithm could then be applied unaltered to the intensity channel, while smoothing of the other two channels could either be performed with the same kernel as in the intensity channel or by isotropic smoothing. The HSV image would then be transformed back to the RGB color space.

However, in very dark color video sequences, there is often a significant difference in the noise levels in the different input channels: for example, the blue channel often has a relatively higher noise level. It is, therefore, essential that it is possible to adapt the algorithm to this difference. To this end, we chose to calculate the structure tensor $\mathbf{J}_\rho$, and its eigenvectors and eigenvalues, in the intensity channel, which we simply define as the mean of the three color channels. The widths of the kernels are then adjusted separately for each color channel by using a different value of the scaling parameter d for each channel. This gives a clear improvement of the output with colors that are closer to the true chromaticity values and with less false color fluctuations than in the aforementioned HSV approach.

When acquiring raw image data from a CCD or complimentary metal–oxide–semiconductor (CMOS) sensor, the pixels are usually arranged according to the so-called Bayer pattern. It has been shown (cf. [25]) that it is efficient and suitable to perform the interpolation from the Bayer pattern to three separate channels, so-called demosaicing, simultaneously to the denoising of the image data. We apply this approach here, for each channel, by setting to zero the coefficients in the summation kernel $k(\mathbf{x_0}, \mathbf{x})$ corresponding to pixels where the intensity data are not available and then normalizing the kernel. A smoothed output is then calculated for both the noisy input pixels and the pixels where data are missing.

V. EXPERIMENTAL RESULTS

In this section, we will describe some experimental results. We have tested our algorithm on a number of different types of sequences, involving static and moving scenes, and with large or little motion of the camera. We have also used a number of different cameras, ranging from cheap consumer cameras to high-end machine vision cameras. We have compared our method to a number of the different methods described in related work section, but most algorithms are not targeted at low-light vision and can not handle the resulting levels of noise. The algorithm that gives comparable results to ours is BM3D for video, [9]. In Fig. 9, the results of running our algorithm are shown for a

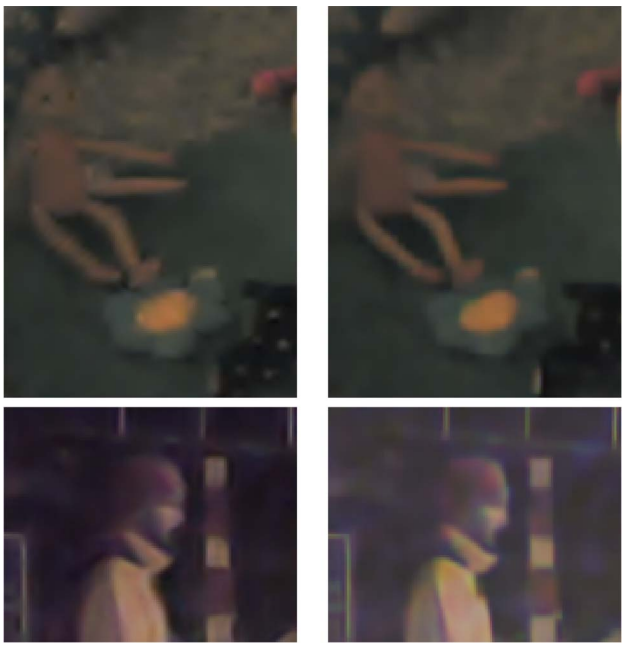

Fig. 10. *Closeup of one frame from the output sequence. To the left is the result using our approach and to the right is the result using BM3D. One can see that BM3D gives very sharp edges but in some cases it hallucinates and removes parts due to the codebook nature of BM3D. One can also see some blocking artefacts due to the block structure of BM3D.*

number of different input sequences. Also shown is the result of running BM3D.

In Fig. 10, a closeup is shown for two of the frames from Fig. 9. One can see that BM3D gives very sharp results with very little noise left. In this sense, the quality is slightly better than of our approach. However, due to the high noise levels two types of artefacts can be noticed in the BM3D images, namely that the block structure is visible, and that we get hallucination effects due to the codebook nature of BM3D. This results in small objects disappearing and appearing.

VI. CONCLUSION

The colors and contrasts of the nocturnal world are just as rich as those found in the diurnal world, and many animals, both vertebrate and invertebrate, have evolved visual systems to exploit this abundant information. Via a combination of highly sensitive optical eye designs, and unique alterations in the morphology, circuitry, and physiology of the retina and higher visual centres, nocturnal animals are capable of advanced and reliable vision at night. In the case of the nocturnal bee *Megalopta genalis*, greatly enlarged corneal facets and rhabdoms, and slow photoreceptors with high contrast gain ensure that visual signal strength is maximal as it leaves the eye and travels to the lamina. Even though it remains to be shown conclusively, anatomical and theoretical evidence suggests that once the visual signals from large groups of ommatidia reach the lamina, they are spatially (and possibly temporally) summed by the second-order monopolar cells, resulting in an enhanced signal and reduced noise. The greatly improved SNR that this strategy could afford, while confined to a narrower range of spatial and temporal frequencies, would ensure that nocturnal visual reliability is maximized for the slower and coarser features of the world. Those features that are faster and finer, and inherently noisy, would be filtered out. But slower and coarser features, in contrast, would be seen more clearly. Using these biological principles as inspiration, we have developed a night-vision algorithm that processes and enhances video sequences captured in very dim light. This algorithm, like a nocturnal visual system, relies on initial amplification of image signals and a local noise-reducing spatio–temporal summation that is weighted by the extent of image motion occurring in the same image locality. The algorithm dramatically increases the visibility of image features, including the preservation of edges, forms, and colors. Not only is this a significant result in the field of dim-light image processing, but also it strengthens the evidence that summation strategies are essential for reliable vision in nocturnal animals. ■

REFERENCES

[1] A. C. Aho, K. Donner, S. Helenius, L. O. Larsen, and T. Reuter, "Visual performance of the toad (*Bufo bufo*) at low light levels, retinal ganglion cell responses and prey-catching accuracy," *J. Compar. Physiol. A*, vol. 172, pp. 671–682, 1993.

[2] A. C. Aho, K. Donner, C. Hydn, L. O. Larsen, and T. Reuter, "Low retinal noise in animals with low body temperature allows high visual sensitivity," *Nature*, vol. 334, pp. 348–350, 1988.

[3] M. Baird, E. Kreiss, W. T. Wcislo, E. J. Warrant, and M. Dacke, "Nocturnal insects use optic flow for flight control," *Biol. Lett.*, vol. 7, pp. 499–501, 2011.

[4] D. Barash and D. Comaniciu, "A common framework for nonlinear diffusion, adaptive smoothing, bilateral filtering and mean shift," *Image Vis. Comput.*, vol. 22, pp. 73–81, 2004.

[5] H. B. Barlow, "Retinal noise and absolute threshold," *J. Opt. Soc. Amer.*, vol. 46, pp. 634–639, 1956.

[6] D. A. Baylor, G. Matthews, and K. W. Yau, "Two components of electrical dark noise in toad retinal rod outer segments," *J. Physiol.*, vol. 309, pp. 591–621, 1980.

[7] E. P. Bennett and L. McMillan, "Video enhancement using per-pixel virtual exposures," in *Proc. SIGGRAPH*, Los Angeles, CA, USA, 2005, pp. 845–852.

[8] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2005, vol. 2, pp. 60–65.

[9] K. Dabov, A. Foi, and K. Egiazarian, "Video denoising by sparse 3D transform-domain collaborative filtering," in *Proc. 15th Eur. Signal Process. Conf.*, 2007, vol. 1, p. 7.

[10] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[11] M. Dacke, E. Baird, M. Byrne, C. Scholtz, and E. J. Warrant, "Dung beetles use the milky way for orientation," *Current Biol.*, vol. 23, pp. 298–300, 2013.

[12] M. Dacke, D. E. Nilsson, C. H. Scholtz, M. Byrne, and E. J. Warrant, "Animal behavior: Insect orientation to polarized moonlight," *Nature*, vol. 424, 2003, DOI: 10.1038/424033a.

[13] F. E. Doujak, "Can a shore crab see a star?" *J. Exp. Biol.*, vol. 166, pp. 385–393, 1985.

[14] F. Drago, K. Myszkowski, T. Annen, and N. Chiba, "Adaptive logarithmic mapping for displaying high contrast scenes," *Proc. EUROGRAPHICS*, vol. 22, pp. 419–426, 2003.

[15] A. Dubs, S. B. Laughlin, and M. V. Srinivasan, "Single photon signals in fly photoreceptors and first order interneurons at behavioural threshold," *J. Physiol.*, vol. 317, pp. 317–334, 1981.

[16] R. Dugad and N. Ahuja, "Video denoising by combining Kalman and Wiener estimates," in *Proc. Int. Conf. Image Process.*, Kobe, Japan, Oct. 1999, vol. 4, pp. 152–156.

[17] R. Frederiksen, W. T. Wcislo, and E. J. Warrant, "Visual reliability and information rate in the retina of a nocturnal bee," *Current Biol.*, vol. 18, pp. 349–353, 2008.

[18] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Reading, MA, USA: Addison-Wesley, 1992.

[19] B. Greiner, W. Ribi, and E. Warrant, "Retinal and optical adaptations for nocturnal vision in the halictid bee *Megalopta genalis*," *Cell Tissue Res.*, vol. 316, no. 3, pp. 377–390, 2004.

[20] B. Greiner, W. A. Ribi, and E. J. Warrant, "A neural network to improve dim-light vision? Dendritic fields of first-order interneurons in the nocturnal bee *Megalopta genalis*," *Cell Tissue Res.*, vol. 323, pp. 313–320, 2005.

[21] B. Greiner, W. A. Ribi, W. T. Wcislo, and E. J. Warrant, "Neuronal organisation in the first optic ganglion of the nocturnal bee *Megalopta genalis*," *Cell Tissue Res.*, vol. 318, pp. 429–437, 2004.

[22] J. H. van Hateren, "Real and optimal neural images in early vision," *Nature*, vol. 360, pp. 68–70, 1992.

[23] J. H. van Hateren, "Spatiotemporal contrast sensitivity of early vision," *Vis. Res.*, vol. 33, pp. 257–267, 1993.

[24] H. Haussecker and H. Spies, "Motion," in *Handbook of Computer Vision and Applications*, vol. 2. New York, NY, USA: Academic, 1999, , pp. 125–151.

[25] K. Hirakawa and T. W. Parks, "Joint demosaicing and denoising," *IEEE Trans. Image Process.*, vol. 15, no. 8, pp. 2146–2157, Aug. 2006.

[26] G. A. Horridge, J. Duniec, and L. Marčelja, "A 24-hour cycle in single locust and mantis photoreceptors," *J. Exp. Biol.*, vol. 91, no. 1, pp. 307–322, 1981.

[27] B. Jähne, *Spatio-Temporal Image Processing*. New York, NY, USA: Springer-Verlag, 1993.

[28] K. Heimonen, I. Salmela, P. Kontiokari, and M. Weckström, "Large functional variability in cockroach photoreceptors: Optimization to low light levels," *J. Neurosci.*, vol. 26, pp. 13454–13462, 2006.

[29] D. S. Kalivas and A. A. Sawchuk, "Motion compensated enhancement of noisy image sequences," in *Proc. IEEE Int. Conf. Acoust. Speech Signal. Process.*, Alburquerque, NM, USA, Apr. 1990, pp. 2121–2124.

[30] S. B. Kang, M. Uytendaele, S. Winder, and R. Szeliski, "High dynamic range video," in *Proc. SIGGRAPH*, 2003, pp. 319–325.

[31] A. Kelber, A. Balkenius, and E. J. Warrant, "Scotopic colour vision in nocturnal hawkmoths," *Nature*, vol. 419, pp. 922–925, 2002.

[32] A. Kelber *et al.*, "Light intensity limits the foraging activity in nocturnal and crepuscular bees," *Behav. Ecol.*, vol. 17, pp. 63–72, 2006.

[33] A. Klaus and E. J. Warrant, "Optimum spatiotemporal receptive fields for vision in dim light," *J. Vis.*, vol. 9, pp. 1–16, 2009.

[34] M. F. Land, "Optics and vision in invertebrates," in *Handbook of Sensory Physiology*, vol. VII/6B, H. Autrum, Ed. New York, NY, USA: Springer-Verlag, 1981, pp. 471–592.

[35] S. B. Laughlin, "Neural principles in the peripheral visual systems of invertebrates," in *Handbook of Sensory Physiology*, vol. VII/6B, H. Autrum, Ed. New York, NY, USA: Springer-Verlag, 1981, pp. 133–280.

[36] S. B. Laughlin, "Invertebrate vision at low luminances," in *Night Vision*, L. T. Sharpe, R. F. Hess, and K. Nordby, Eds. Cambridge, U.K.: Cambridge Univ. Press, 1990, pp. 223–250.

[37] S. B. Laughlin, "Matched filtering by a photoreceptor membrane," *Vis. Res.*, vol. 36, pp. 1529–1541, 1996.

[38] S. B. Laughlin, A. D. Blest, and S. Stowe, "The sensitivity of receptors in the posterior median eye of the nocturnal spider *Dinopis*," *J. Compar. Physiol.*, vol. 141, pp. 53–65, 1980.

[39] S. B. Laughlin and P. G. Lillywhite, "Intrinsic noise in locust photoreceptors," *J. Physiol.*, vol. 332, pp. 25–45, 1982.

[40] S. B. Laughlin and M. Weckström, "Fast and slow photoreceptors: A comparative study of the functional diversity of coding and conductances in the Diptera," *J. Compar. Physiol. A*, vol. 172, pp. 593–609, 1993.

[41] S.-W. Lee, V. Maik, J. Jang, J. Shin, and J. Paik, "Noise-adaptive spatio-temporal filter for real-time noise removal in low light level images," *IEEE Trans. Consumer Electron.*, vol. 51, no. 2, pp. 648–653, May 2005.

[42] S. H. Lee and M. G. Kang, "Spatio-temporal video filtering algorithm based on 3-D anisotropic diffusion equation," in *Proc. Int. Conf. Image Process.*, 1998, vol. 2, pp. 447–450.

[43] P. G. Lillywhite, "Single photon signals and transduction in an insect eye," *J. Compar. Physiol.*, vol. 122, pp. 189–200, 1977.

[44] P. G. Lillywhite and S. B. Laughlin, "Transducer noise in a photoreceptor," *Nature*, vol. 277, pp. 569–572, 1979.

[45] H. Malm and E. Warrant, "Motion dependent spatiotemporal smoothing for noise reduction in very dim light image sequences," in *Proc. Int. Conf. Pattern Recognit.*, Hong Kong, Sep. 2006, pp. 954–959.

[46] H. Malm, M. Oskarsson, and E. Warrant, "Biologically inspired enhancement of dim light video," in *Frontiers in Sensing*. New York, NY, USA: Springer-Verlag, 2012, pp. 71–85.

[47] H. Malm *et al.*, "Adaptive enhancement and noise reduction in very low light-level video," in *Proc. IEEE 11th Int. Conf. Comput. Vis.*, 2007, DOI: 10.1109/ICCV.2007.4409007.

[48] G. R. Martin, *Birds by Night*. London, U.K.: Poyser, 1990.

[49] K. Miyata and A. Taguchi, "Spatio-temporal separable data-dependent weighted average filtering for restoration of the image sequences," in *Proc. IEEE Int. Conf. Acoust. Speech. Signal. Process.*, Alburquerque, NM, USA, May 2002, vol. 4, pp. 3696–3699.

[50] J. E. Niven, J. C. Anderson, and S. B. Laughlin, "Fly photoreceptors demonstrate energy-information trade-offs in neural coding," *PLoS Biol.*, vol. 5, no. 4, 2007, e116.

[51] J. E. Niven *et al.*, "The contribution of shaker k+ channels to the information capacity of *Drosophila* photoreceptors," *Nature*, vol. 421, no. 6923, pp. 630–634, 2003.

[52] K. P. Ohly, "The neurons of the first synaptic regions of the optic neuropil of the firefly *Phausius splendidula L.* (Coleoptera)," *Cell Tissue Res.*, vol. 158, pp. 89–109, 1975.

[53] M. K. Özkan, M. I. Sezan, and A. M. Tekalp, "Adaptive motion-compensated filtering of noisy image sequences," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 3, no. 4, pp. 277–290, Aug. 1993.

[54] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629–639, Jul. 1990.

[55] K. Pirhofer-Walzl, E. J. Warrant, and F. G. Barth, "Adaptations for vision in dim light: Impulse responses and bumps in nocturnal spider photoreceptor cells (*Cupiennius salei* keys)," *J. Compar. Physiol. A*, vol. 193, pp. 1081–1087, 2007.

[56] S. M. Pizer *et al.*, "Adaptive histogram equalization and its variations," *Comput. Vis. Graph. Image Process.*, vol. 39, pp. 355–368, 1987.

[57] S. F. Reid, A. Narendra, J. M. Hemmi, and J. Zeil, "Polarised skylight and the landmark panorama provide night-active bull ants with compass information during route following," *J. Exp. Biol.*, vol. 214, pp. 363–370, 2011.

[58] E. Reinhard, G. Ward, S. Pattanaik, and P. Debevec, *High Dynamic Range Imaging: Acquisition, Display, Image-Based Lighting*. San Mateo, CA, USA: Morgan Kaufmann, 2005.

[59] W. A. Ribi, "Fine structure of the first optic ganglion (lamina) of the cockroach *Periplaneta americana*," *Tissue Cell*, vol. 9, pp. 57–72, 1977.

[60] W. A. Ribi, "The first optic ganglion of the bee," *Cell Tissue Res.*, vol. 165, no. 1, pp. 103–111, 1975.

[61] A. Rose, "The relative sensitivities of television pickup tubes, photographic film and the human eye," *Proc. Inst. Radio Eng. New York*, vol. 30, pp. 293–300, 1942.

[62] M. I. Sezan, M. K. Özkan, and S. V. Fogel, "Temporally adaptive filtering of noisy image sequences," in *Proc. IEEE Int. Conf. Acoust. Speech Signal. Process.*, Toronto, ON, Canada, May 1991, vol. 4, pp. 2429–2432.

[63] A. W. Snyder, "Acuity of compound eyes, physical limitations and design," *J. Compar. Physiol.*, vol. 116, pp. 161–182, 1977.

[64] A. W. Snyder, S. B. Laughlin, and D. G. Stavenga, "Information capacity of eyes," *Vis. Res.*, vol. 17, pp. 1163–1175, 1977.

[65] A. W. Snyder, D. G. Stavenga, and S. B. Laughlin, "Spatial information capacity of compound eyes," *J. Compar. Physiol.*, vol. 116, pp. 183–207, 1977.

[66] H. Somanathan, R. M. Borges, E. J. Warrant, and A. Kelber, "Nocturnal bees learn landmark colours in starlight," *Current Biol.*, vol. 18, 2008. [Online]. Available: http://dx.doi.org/10.1016/j.cub.2008.08.023

[67] H. Somanathan, R. M. Borges, E. J. Warrant, and A. Kelber, "Visual ecology of Indian carpenter bees I: Light intensities and flight activity," *J. Compar. Physiol. A*, vol. 194, pp. 97–107, 2008.

[68] D. G. Stavenga and J. W. Kuiper, "Insect pupil mechanisms. i. on the pigment migration in the retinula cells of *Hymenoptera* (suborder Apocrita)," *J. Compar. Physiol.*, vol. 113, no. 1, pp. 55–72, 1977.

[69] N. J. Strausfeld and A. D. Blest, "Golgi studies on insects I. The optic lobes of *Lepidoptera*," *Philosoph. Trans. Roy. Soc. Lond. B*, vol. 258, pp. 81–134, 1970.

[70] J. Theobald, E. J. Warrant, and D. C. O'Carroll, "Wide-field motion tuning in nocturnal hawkmoths," *Proc. Roy. Soc. Lond. B*, vol. 277, pp. 853–860, 2010.

[71] J. C. Theobald, B. Greiner, W. T. Wcislo, and E. J. Warrant, "Visual summation in night-flying sweat bees: A theoretical study," *Vis. Res.*, vol. 46, pp. 2298–2309, 2006.

[72] S. M. Tierney, O. Sanjur, G. G. Grajales, L. M. Santos, E. Bermingham, and W. T. Wcislo, "Photic niche invasions: Phylogenetic history of the dim-light foraging augochlorine bees (*Halictidae*)," *Proc. Roy. Soc. B, Biol. Sci.*, vol. 279, no. 1729, pp. 794–803, 2012.

[73] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. 6th Int. Conf. Comput. Vis.*, 1998, pp. 839–846.

[74] D. Uttenweiler, C. Weber, B. Jähne, R. H. A. Fink, and H. Scharr, "Spatiotemporal anisotropic diffusion filtering to improve signal-to-noise ratios and object restoration in fluorescence microscopic image sequences," *J. Biomed. Opt.*, vol. 8, no. 1, pp. 40–47, 2003.

[75] De Vries and H. , "The quantum character of light and its bearing upon threshold of vision, the differential sensitivity and visual acuity of the eye," *Physica*, vol. 10, pp. 553–564, 1943.

[76] E. J. Warrant, "Seeing better at night: Life style, eye design and the optimum strategy of spatial and temporal summation," *Vis. Res.*, vol. 39, pp. 1611–1630, 1999.

[77] E. J. Warrant, "Vision in the dimmest habitats on earth," *J. Compar. Physiol. A*, vol. 190, pp. 765–789, 2004.

[78] E. J. Warrant, "The sensitivity of invertebrate eyes to light," in *Invertebrate Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2006, pp. 167–210.

[79] E. J. Warrant, "Nocturnal vision," in *The Senses: A Comprehensive Reference (Vol. 2: Vision II)*, T. Albright and R. H. Masland, Eds. New York, NY, USA: Academic, 2008, pp. 53–86.

[80] E. J. Warrant, "Seeing in the dark: Vision and visual behaviour in nocturnal bees and wasps," *J. Exp. Biol.*, vol. 211, pp. 1737–1746, 2008.

[81] E. J. Warrant and M. Dacke, "Vision and visual navigation in nocturnal insects," *Annu. Rev. Entomol.*, vol. 56, pp. 239–254, 2011.

[82] E. J. Warrant, T. Porombka, and W. H. Kirchner, "Neural image enhancement

allows honeybees to see at night," *Proc. Roy. Soc. Lond. B*, vol. 263, pp. 1521–1526, 1996.

[83] W. T. Wcislo, "The evolution of nocturnal behaviour in sweat bees *Megalopta genalis* and *M ecuadoria* (Hymenoptera, Halictidae): An escape from competitors and enemies?" *Biol. J. Linnean Soc.*, vol. 83, pp. 377–387, 2004.

[84] W. T. Wcislo and S. M. Tierney, "Behavioural environments and niche construction: The evolution of dim-light foraging in bees," *Biol. Rev.*, vol. 84, pp. 19–37, 2009.

[85] M. Weckström and S. B. Laughlin, "Visual ecology and voltage-gated ion channels in insect photoreceptors," *Trends Neurosci.*, vol. 18, no. 1, pp. 17–21, 1995.

[86] J. Weickert, *Anisotropic Diffusion in Image Processing*. Stuttgart, Germany: Teubner-Verlag, 1998.

[87] J. Weickert, "Coherence-enhancing diffusion filtering," *Int. J. Comput. Vis.*, vol. 31, no. 2–3, pp. 111–127, 1999.

[88] G. Z. Yang, P. Burger, D. N. Firmin, and S. R. Underwood, "Structure adaptive anisotropic image filtering," *Image Vis. Comput.*, vol. 14, pp. 135–145, 1996.

ABOUT THE AUTHORS

Eric Warrant received an honors degree in physics from the University of New South Wales, Sydney, N.S.W., Australia, in 1985 and the Ph.D. degree in visual science from the Australian National University, Canberra, A.C.T., Australia, in 1990, studying the designs and optics of insect eyes.

He is a Professor of Zoology at the University of Lund, Lund, Sweden. He took up a Postdoctoral position at Lund University in 1990, became tenured in 1996, and became a full professor in 2002. He leads an active research group studying vision and visual navigation in animals from extremely dim habitats (nocturnal and deep sea). He has published two books, and over 100 original papers, reviews, and popular articles, and his work has frequently been featured in the national and international press.

Prof. Warrant is Vice Chairman of the National Committee for Biology (Royal Swedish Academy of Sciences). He is a Fellow and Board Member of the Royal Physiographic Society and a Foreign Fellow of the Royal Danish Academy of Sciences and Letters.

Magnus Oskarsson (Member, IEEE) received the M.Sc degree in engineering physics and the Ph.D. degree in mathematics from the University of Lund, Lund, Sweden, in 1997 and 2002, respectively. His thesis work was devoted to computer vision with applications for autonomous vehicles.

He is currently an Associate Professor at the Centre for Mathematical Sciences, Lund University, where his teachings include undergraduate and graduate courses in mathematics and image analysis. He is the author and coauthor of a number of papers published in international journals and conference proceedings within geometry, algebra, and optimization with applications in computer vision, cognitive vision, and image enhancement.

Henrik Malm received the M.Sc. degree in computer science and technology and the Ph.D. degree in applied mathematics from Lund Institute of Technology, Lund University, Lund, Sweden, in 1998 and 2003, respectively, and the B.Sc. degree in architecture from Lund Institute of Technology in 2012.

In autumn 2000, he was a Guest Researcher with the Computer Vision Laboratory, Center for Automation Research, University of Maryland, College Park, MD, USA. He was a Postdoctoral Fellow with the Lund Vision Group at the Department of Cell and Organism Biology, Lund University. He has authored a number of papers on the following subjects in international journals and conference proceedings: robot vision, camera calibration, visual illusions, nonlinear diffusion filtering, and biologically inspired night-vision systems. He is currently working as a practicing architect at the architecture practice Fosters and Partners in London (UK) and as a researcher and lecturer in architecture, specializing on aspects of computational design.