ORIGINAL PAPER

Biological
Cybernetics

Bio-inspired visual ego-rotation sensor for MAVs

**Johannes Plett · Armin Bahl · Martin Buss ·
Kolja Kühnlenz · Alexander Borst**

Abstract Flies are capable of extraordinary flight maneu-
vers at very high speeds largely due to their highly elaborate
visual system. In this work we present a fly-inspired FPGA
based sensor system able to visually sense rotations around
different body axes, for use on board micro aerial vehicles
(MAVs). Rotation sensing is performed analogously to the
fly's VS cell network using zero-crossing detection. An addi-
tional key feature of our system is the ease of adding new
functionalities akin to the different tasks attributed to the fly's
lobula plate tangential cell network, such as object avoidance
or collision detection. Our implementation consists of a mod-
ified eneo SC-MVC01 SmartCam module and a custom built
circuit board, weighing less than 200 g and consuming less
than 4 W while featuring 57,600 individual two-dimensional
elementary motion detectors, a 185° field of view and a frame
rate of 350 frames per second. This makes our sensor system
compact in terms of size, weight and power requirements for
easy incorporation into MAV platforms, while autonomously
performing all sensing and processing on-board and in real
time.

Keywords Reichardt Detector · Rotation estimation ·
Motion detection · MAV · Fly vision · Biorobotics

J. Plett (✉) · A. Bahl · A. Borst
Department of Systems and Computational Neurobiology,
Max Planck Institute of Neurobiology, 82152 Martinsried,
Germany
e-mail: jplett@neuro.mpg.de

M. Buss · K. Kühnlenz
Institute of Automatic Control Engineering (LSR), Technische
Universität München (TUM), 80290 Munich, Germany

1 Introduction

Perception of visual motion has been an intense and fruit-
ful field of research over many decades. Especially studies
of insects—and flies in particular—have revealed astound-
ingly simple, yet robust and elegant solutions of extracting
motion information from noisy and complex environments.
Flies are able to autonomously navigate at very high speeds
through highly unstructured settings, by and large relying
only on visual cues. Despite having only a few 100,000
neurons, they are able to achieve these feats because of the
highly optimized way these neurons are interconnected and
the ideally suited basic operation principles of motion vision.
Flies extract cues about motion relative to the environment
from the *optic flow* at remarkably high temporal resolution.
The true optic flow is the velocity field of the projection of
the relative motion between observer and visual surround-
ings onto the retina. Given that this true optic flow is not
directly measurable it is estimated from spatiotemporal lumi-
nance patterns on the retina by dedicated neuronal circuits.
Since these dedicated circuits are very effective, robust, and
efficient in terms of implementation they lend themselves
well for technical applications.

In recent years, unmanned aerial vehicles (UAVs) and
micro aerial vehicles (MAVs) have become more and more
common in tasks, such as aerial reconnaissance, surveillance,
and exploration. To cope with the rising complexity of these
challenges increasing levels of automation are needed. This
usually leads to larger and computationally more intense
solutions which require large on-board processing units (e.g.,
Franceschini et al. 1992) somewhat limiting their use on
board small flying vehicles. One solution to this problem is
"out-sourcing" of computational load to off-board comput-
ing platforms (e.g., Bermudez i Badia et al. 2007; Kendoul
et al. 2009; Zhang et al. 2008). This, however, is often not

possible due to inadequacies of wireless transmission, such as low throughput, large delays, jitter, temporary loss of signal, etc. A promising way of solving these issues is the onboard use of highly efficient algorithms, such as those found in biological vision systems. In fact, over the past decades the insect visual system has inspired many studies towards visually guided autonomous vehicles. Much emphasis has been put on the implementation of collision avoidance strategies (e.g., Harrison 2005; Bermudez i Badia et al. 2007) and local navigation (e.g., Zufferey and Floreano 2006; Srinivasan et al. 2009; Conroy et al. 2009; Moeckel and Liu 2009; Beyeler et al. 2009; Hyslop et al. 2010). Moreover, considerable work has been put forth on autonomous height control (e.g., Netter and Franceschini 2002; Valette et al. 2010).

One aspect of fly motion vision that has received relatively little attention in technical implementations is rotation sensing. There have been studies on basic motion detection circuits for rotation detection (O'Carroll et al. 2006; Aubepart et al. 2004), but despite considerable advances in understanding of the fly neuronal rotation sensing circuitry (Krapp et al. 1998; Borst et al. 2010; Cuntz et al. 2007) there have been few biologically realistic practical applications involving these findings. O'Carroll et al. (2006) have put forth a rotation sensor using a custom aVLSI chip that relies on basic motion detection circuitry for a one-dimensional circular array of 40 input photodiodes. Aubepart et al. (2004) used a Field Programmable Gate Array (FPGA) based solution with a linear 12-photodiode array, theoretically capable of handling up to 245 input elements. Köhler et al. (2009) proposed a solution of higher spatial resolution at 120×100 input pixels over a 40° horizontal field of view and a temporal resolution of 100 frames per second (fps), expandable up to 200 fps in bright outdoor conditions. But despite promising results in artificially structured environments, the system did not work in naturalistic settings. A similarly oriented approach was used by Zhang et al. (2008). They successfully implemented 256×256 motion detection circuits operating at 350 fps and six motion templates for template matching based motion detection. However, their system architecture residing on a PCI-FPGA card in a host PC forfeits use on board small aerial vehicles. Also, to the authors' best knowledge there is currently no commercially available visual ego-rotation sensor for this specific purpose.

In this study we set out to implement a small and lightweight fly-inspired visual rotation sensor for MAVs, keeping algorithmically as close as possible to the biological model while maintaining similar spatial and temporal resolution over a similar field of view.

2 Fly motion vision

The fly motion vision system can be segmented into several distinct functional and anatomical units. The input layer is the compound eye, which consists of a hexagonal array of several hundreds to thousands of *ommatidia*, each harboring a lenslet and a set of photoreceptor cells. This stage constitutes the retina, from where information is passed retinotopically on to three successive neuropiles, the lamina, medulla, and lobula complex. In the medulla, local motion estimates are computed according to the detector model put forward by Hassenstein and Reichardt (1956), commonly known as the Reichardt Detector or elementary motion detector (EMD). As depicted in Fig. 1a, the simplest form of the Reichardt Detector consists of two mirror-symmetric subunits, each correlating two spatially adjacent input signals with each other by multiplying one input signal with a temporally low-pass filtered version of the other. The output of both subunits is then subtracted, yielding a direction-selective output while suppressing non-motion artifacts. This way of estimating motion is particularly well suited for applications in presence of noise, i.e., with poor signal to noise ratio (Potters and Bialek 1994; Borst 2007). However, it is not a perfect velocity estimator as it depends not only on velocity but also on local texture and contrast (Reichardt and Egelhaaf 1988; Egelhaaf et al. 1989). Furthermore, individual local motion estimators suffer from the *aperture problem* due to their limited field of view (Stumpf 1911).

To circumvent these problems flies spatially integrate local motion estimates over larger areas, thus to a large extent averaging out the aforementioned effects (Single and Borst 1988). This is done in the lobula plate by large interneurons called lobula plate tangential cells (LPTCs). These neurons form an ensemble of roughly 60 uniquely identifiable cells, out of which two prominent groups—the vertical system (VS) and horizontal system (HS) cells—are preferentially sensitive to vertical and horizontal motion, respectively. Per hemisphere, the blowfly *Calliphora erythrocephala* possesses ten VS cells VS 1 through VS 10, whose dendritic receptive fields sequentially cover narrow but overlapping vertical stripes of the visual field, going around the dorsoventral axis from frontal (VS 1) to caudal (VS 10). Each VS cell integrates the responses from local vertical motion detectors within its own specific receptive field. Strikingly, the response of VS cells in their axon terminal regions suggest much broader receptive fields (Elyada et al. 2009). This broadening of the axon terminal response has been shown to be caused by gap junctions interconnecting the VS cells (Haag and Borst 2004, 2005; Farrow et al. 2005). Furthermore, VS 1 and VS 10 cells mutually inhibit each other (Haag and Borst 2004, 2007). This gives rise to the VS cell network illustrated in Fig. 2a with its associated connection strength matrix given in Fig. 2b. The reason for this network scheme is thought to be strengthening of robustness to inhomogeneities of pattern contrast, i.e., making this system more suitable for use in naturalistic environments (Cuntz et al. 2007; Elyada et al. 2009; Wertz et al. 2009). We note that the model

Fig. 1 **a** Basic Reichardt
Detector consisting of two
mirror-symmetric units, each
correlating one input with a
low-pass filtered version of the
other. Final subtraction of both
sub-unit outputs yields a
directionally selective signal
depending on stimulus velocity
and direction.
b Two-dimensional arrangement
of Reichardt Detectors giving
rise to 2D local motion
estimates by estimating vertical
and horizontal components of
local motion. **c** Elaborated
Reichardt Detector including
automatic gain adaptation and a
homomorphic filtering stage
sequentially connected giving
input to a basic Reichardt
Detector. **d** VS and HSN cell
dendritic field distribution over
the half sphere captured by the
185° fisheye lens showing
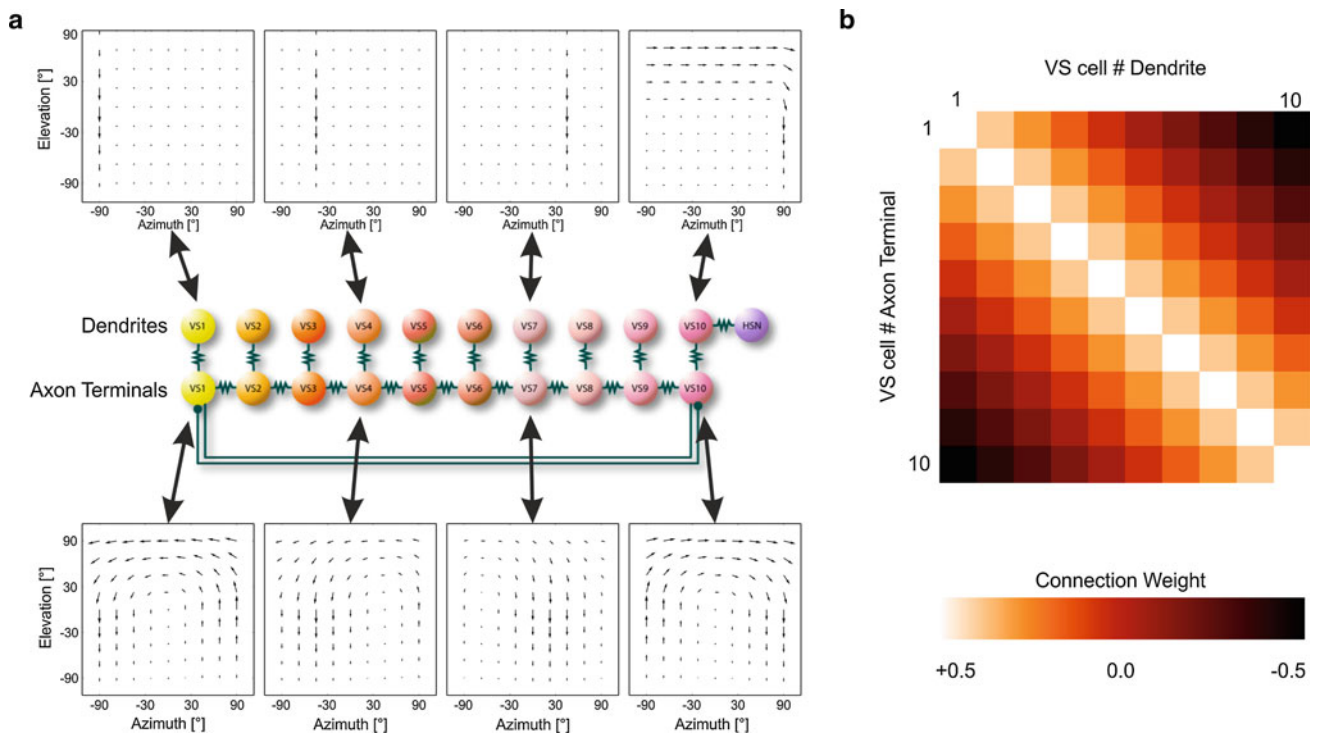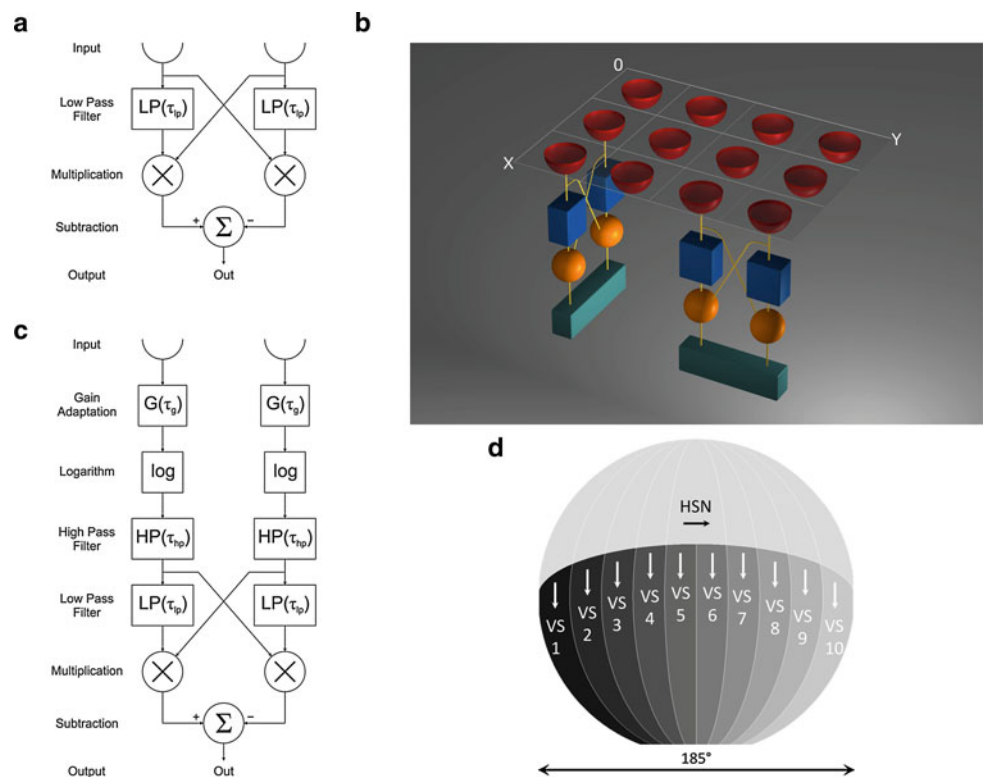direction of motion sensitivity
of each cell

Fig. 2 VS cell network coupling. **a** VS cell network showing coupling between adjacent cells in their axon terminal regions, including bi-directional mutual inhibition between VS 1 and VS 10. HSN and VS 10 are connected directly in the dendritic regions. Ideal receptive fields of individual cells are shown for both the dendritic and the axon terminal regions. Whereas receptive fields in the dendritic regions exhibit but one narrow vertical stripe, the receptive fields in the axon terminals regions resemble matched filters for rotations around different axes along the equator. **b** Connection strength matrix for network coupling. The axon terminal output of a cell is exited strongly by its own dendritic input, slightly less by its immediate neighbors, little by more distal cells and even inhibited by most distal cells

analyzed here additionally includes the effect of the HSN cell on the dendritic compartment of VS 10, thereby accounting for reported responses to dorsal horizontal motion (Krapp et al. 1998; Haag and Borst 2003, 2007).

As proposed by Cuntz et al. (2007) the responses of the VS cell network in its axonally coupled form can be used to robustly infer the approximate center of rotation for rotations around axes lying on the equatorial plane. Due to their vertical directional sensitivity, the VS cells on one side of the center of rotation will respond by hyperpolarizing, while to the other side they will depolarize. The location where VS cell responses change signs, i.e., the *zero-crossing* location, indicates the approximate location of the center of rotation.

3 System design

3.1 Requirements and restrictions

The goal of this study was the construction of an optic flow based sensor system that is algorithmically as close as possible to the biological original of the fly visual system, in particular the VS cell network. Therefore the design requirements included similar spatiotemporal resolution and field of view (FOV) compared to a fly, as well as reasonable light weight, low power consumption, and compact size.

The blowfly *Calliphora* is able to detect flicker up to rates between 200 and 300 Hz (Autrum 1952) or even higher (Tatler et al. 2000). Thus, the design goal was set to achieve frame rates well above 300 fps, exceeding cutoff frequencies of 150 Hz. Each compound eye of *Calliphora* extends about 190° in the horizontal and 198° in the vertical plane (Seitz 1968). To achieve a large FOV the camera system was equipped with a fisheye lens covering a solid angle of approximately 2π sr, i.e., half of the unit sphere. The highest spatial resolution found in *Calliphora* amounts to inter-ommatidial angles of $\Delta\varphi = 1.07°$ (Land and Eckert 1985) and is reached in the frontal visual field (Petrowitz et al. 2000). Thus, to obtain a spatial resolution better than 1° per pixel in the frontal part while using a 185° fisheye lens the sensor system had to have a resolution of at least 185×185 pixels. Commonly found MAVs are able to carry payloads only up to a few hundred grams. The envisaged primary test platform for this sensor system, the AscTec Hummingbird quadrocopter (Ascending Technologies, Krailling, Germany), features a payload of up to 200 g. Hence, the weight restrictions of the sensor system were fixed to an upper limit of 200 g. Due to these weight restrictions, battery power on board is limited. We specified power consumption restrictions to a maximum of 4 W. In terms of size the system was required to be able to be mounted on such an MAV without interfering much with its aerodynamics.

3.2 Computations

The computation of rotational axis and velocity estimates was divided into the following five pipelined sequential steps:

Image acquisition is done by the image sensor in a row-wise fashion pixel by pixel at full frame rate,
Pre-processing suppresses illumination artifacts by automatic gain adaptation and homomorphic filtering,
Local motion detection is performed using a Reichardt Detector correlation model,
Global motion integration is achieved by wide-field integration of local motion estimates,
Rotation estimation is accomplished by calculating location and slope of the VS cell network zero-crossing.

The individual processing steps are described in more detail in the following sections.

3.2.1 Pre-processing

At the core of this sensor design lies the aforementioned Reichardt Detector or EMD. An EMD inherently displays a quadratic dependence on image contrast which makes it also sensitive to changes in overall lighting. To improve robustness against lighting changes a homomorphic filtering approach (Gonzalez and Woods 2007) was applied as a pre-processing stage to the EMD. In a visual scene, illumination and reflectance combine multiplicatively and are therefore not linearly separable. Nevertheless, they usually occupy distinct regions in the frequency domain since illumination tends to vary slowly in time and space while reflectance provides mostly high temporal frequency components due to reflections from objects. For a given pixel in an image its value is given by

$$I(x, y, t) = L(x, y, t) \cdot R(x, y, t) \qquad (1)$$

where $I(x, y, t)$ represents the value of the pixel at location (x, y) at time point t, while $L(x, y, t)$ and $R(x, y, t)$ represent illumination and reflectance for that location and time point. By taking the logarithm of the pixel value these two components become additive (Eq. 2) and the low frequency illumination components can be filtered out using a high-pass filter, leaving only reflectance (Eq. 3).

$$\log(I(x, y, t)) = \log(L(x, y, t)) + \log(R(x, y, t)) \qquad (2)$$

$$\mathrm{HP}(\log(I(x, y, t))) \approx \log(R(x, y, t)) \qquad (3)$$

Using this homomorphic filtering technique the elaborated Reichardt Detector used for final implementation effectively included a logarithmic stage via lookup table and a first-order

temporal high-pass filter acting together as an input stage to a basic EMD (Fig. 1c).

To further optimize the dynamic range of the image sensor pixel values over a large illumination range an automatic camera gain adaptation control was implemented. The temporally low-pass filtered mean pixel values of each frame ($\tau_g = 1s$) were utilized as a crude measure for overall illumination. A simple proportional controller was used to adjust the internal camera gain as to keep the mean pixel values reasonably centered within the sensor coding range.

3.2.2 Local and global motion detection

For local motion detection of each pixel the elementary Reichardt Detector of Fig. 1a was used in conjunction with a homomorphic pre-processing stage, constituting the elaborated EMD (Fig. 1c). Each incoming pixel value is thusly correlated with its immediate horizontal and vertical neighbor, giving rise to a two-dimensional local motion estimate (Fig. 1b).

For wide field integration of these local motion estimates a network akin to the Calliphora VS cell network was established. For all ten VS cell homologues the vertical components of local motion estimates within their respective receptive fields are linearly summed up. For the three HS cell homologues this was done for the respective horizontal components. The ten VS cells' receptive fields are linearly spaced along the equator, each covering a tenth of the fish-eye lens projection on the image plane, ranging from VS 1 on the far left up to VS 10 on the far right (Fig. 1d). Similarly, the HSN receptive field covers the upper third of the lens projection (Fig. 1d), the HSE receptive field the middle third and the HSS receptive field the lower third. The HSE and HSS cells, however, were not used for rotation sensing and therefore included for future extensions only.

To improve robustness as in the biological original (Cuntz et al. 2007) the cells in the network were interconnected as outlined in Fig. 2. In this wiring scheme adjacent cells are strongly coupled while most distant cells are mutually inhibitory, as indicated by the connection matrix of Fig. 2b. This yielded a robust and symmetrical response pattern of the network.

3.2.3 Rotation estimation

Estimation of the axis of rotation based on the VS cell output relies on the fact that VS cells' receptive fields resemble matched filters for rotations around rotations sequentially arranged around the dorso-ventral axis on the equatorial plane (Krapp et al. 1998). As introduced in Sect. 2, by calculating the zero-crossing location of the VS cell network responses the center of rotation can be inferred. Furthermore, at that location the slope of the curve is strongly correlated

with the rate and direction of rotation. If the curve has a positive slope going from a negative VS 1 response to a positive VS 10 response the rotation of the visual scene is clockwise. Accordingly, a negative slope indicates a counter-clockwise rotation. Also, a fast rotation would produce a steep slope, whereas slower rotations would yield a more shallow slope. Hence, this slope magnitude correlates directly with the rate of rotation, albeit in a nonlinear bell-shaped fashion.

3.3 Implementation

3.3.1 Components overview

The key challenges of the implementation of this sensor system were the high computation rate and small footprint required. For the computing platform the typical choices were off-board computation on a PC or on-board computation using microcontrollers, microprocessors, programmable logic, or fully custom designed chips. For the off-board computation the images would have to be first sent from the MAV to the PC, which with the current wireless transmission standards, such as WLAN, ZigBee, or Bluetooth is not yet possible at frame rates much higher than around 100 fps. Therefore, the choices were limited to on-board solutions, of which due to the high throughput requirements and size/weight constraints microcontrollers and sequential general purpose microprocessors were ruled out. Fully custom designed Application Specific Integrated Circuits (ASICs) were not an option due to their high cost and time consuming design cycles. Since optic flow calculations are highly parallel FPGAs were the ideal choice owing to their inherently parallel and pipelineable nature, thus permitting high throughputs.

As the core image capture and processing unit an eneo SmartCam SC-MVC01 module (Videor, Rödermark, Germany) was chosen, being able to provide a spatial resolution of 640 × 240 pixels at 370 fps and 8 bit resolution in row interlaced mode. It features a 1/2 inch Micron MT9V403 CMOS image sensor, an Intel XScale PXA255 processor running at 400 MHz, a Xilinx Spartan-3 series XC3S1000 FPGA and an Infineon HYB25L256160AF 256 Mbit Mobile-RAM module accessible freely from the FPGA. Using a 185° DSL215B-NIR miniature fisheye lens (Sunex, Carlsbad, USA) and a custom built light weight camera backplane the camera possessed a 72 mm × 45 mm × 45 mm footprint weighing 148 g. Additional processing was carried out on an Atmel AT91SAM7A3 ARM processor (Atmel, San Jose, USA) on a custom designed printed circuit board (PCB) also housing power conversion circuitry and communication interfaces. Overall hardware costs amounted to approximately €2,000.

In order to monitor the outcome of the optic flow calculation, ego motion estimation, and the captured camera images,
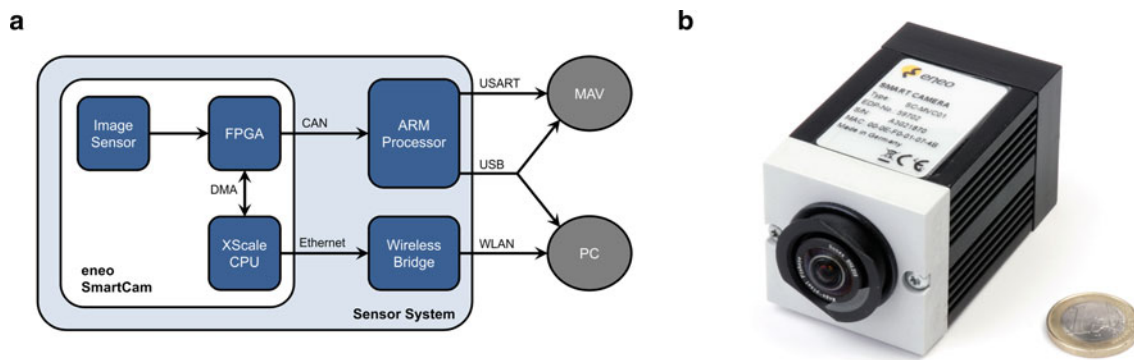
Fig. 3 **a** General system architecture. At the system core lies an eneo SmartCam with its embedded FPGA and XScale processor, weighing 148 g including the fisheye lens and consuming 1.69 W. Further processing and communication with the MAV and an optional real-time data acquisition PC is carried out using an ARM processor hosted on a custom PCB weighing 30 g and consuming 0.51 W. For wireless transmission of live images and processed data towards the PC ground station the eneo SmartCam Ethernet port is used via a wireless bridge weighing 15 g and consuming 1.47 W. Total system weight was 193 g while consuming a total of 3.67 W. **b** eneo SmartCam including 185° fisheye lens in size comparison with a €1 coin

a wireless communication system was used to communicate with an external laptop PC (Dell, Round Rock, USA) hosting a control and monitoring interface. A general schematic of the system hardware architecture is given in Fig. 3.

3.3.2 FPGA design

Pre-processing, computation of the local motion estimates, and subsequent spatial integration was implemented on board the Xilinx Spartan-3 FPGA using VHDL. The internal design was broken down into several modules dealing with specific tasks, such as image data acquisition from the image sensor, local motion estimation, external SDRAM communication and management, large field integration of local motion estimates, internal timing management, communication with the off-board ARM processor and monitoring PC, etc.

The fisheye lens projects a centered circular image onto the imager's row interlaced 640×240 pixels, but only the central 240×240 pixels were used. To estimate local motion, each pixel was correlated with the adjacent left and upper pixel, using the elaborated EMD, resulting in a two-dimensional local motion vector for each pixel. The distance between EMD input arms $\Delta\varphi$—equivalent to the fly's inter-ommatidial angle—is thus equal to 1 pixel, which in the frontal part of the the FOV equates to $\Delta\varphi = \frac{185°}{240} = 0.77°$.

Taking advantage of an FPGAs inherent parallel capabilities, the local motion estimate computation was implemented in a pipelined fashion, thus, reducing the elaborated EMD to 15 atomic instructions (such as memory fetch, table look up, multiplication, sum, and subtraction operations), each being executed in strictly less than 20 ns. For multiplication, dedicated hardware multipliers of the Spartan-3 series were used. EMD computations were carried out at 18 bit Q10.7 fixed point precision, thereby accounting for fractional results ensuring minimal loss of precision through truncation.

The elaborated EMD incorporates temporal low-pass and high-pass filters whose immediate results need to be stored between image cycles, yielding an amount of data of over four times the total internal Block RAM storage capacity of the Spartan-3 XC3S1000. Therefore, the external SDRAM attached to the FPGA needed to be used to store and retrieve inter-frame filter data. An SDRAM controller module loosely based on application notes by XILINX (1999, 2003) was implemented, operating in half-duplex mode at 100 MHz and a 16 bit data bus width.

A wide field integration module was written to calculate the dendritic part of the VS cell network output from the local motion estimates via Boolean map lookup, yielding one scalar value for each VS and HS cell.

Also a communication module was implemented for relaying the resulting wide field integration data at full frame rate towards the external ARM processor. The SmartCam hardware was modified in a way that its High Speed CAN bus output could be used directly by the FPGA. For establishing the communication with the ARM processor a custom CAN bus controller FPGA core operating at 1 Mbit/s was written. The communications module also handles data transfers from and towards the SmartCam internal Intel XScale processor via the shared 64 MByte SDRAM memory between FPGA and XScale processor using Direct Memory Access (DMA).

Using a speed optimized XST synthesizer the complete design occupied 47% of available slice flip flops, 69% of all 4-input look up tables (LUTs) and 83% of available block RAM of the Spartan-3 XC3S1000.

3.3.3 XScale firmware

The internal Intel XScale processor of the SmartCam module controls several variables and parameters of the image sensor, such as operation modes, buffer sizes, and frame rate. These

parameters along with incoming image data are transferred via DMA through the shared SDRAM memory. For communication towards an external PC the SmartCam features a 10/100 Mbit/s Ethernet MAC/PHY directly connected to the processor. Its operating system is an embedded Linux Kernel 2.6.6 for which a resident camera daemon application was written in C++ that takes care of initialization routines, handshaking protocols, and communication between XScale processor and FPGA as well as between XScale processor and the ground station PC. To communicate wirelessly between camera system and ground station PC the internal PCB of an Asus WL-330gE wireless bridge (Asus, Taipei, Taiwan) was used. One transmitted data frame consists of an image, local motion estimates, and the associated wide field integration results.

3.3.4 ARM firmware

For further processing of the raw wide field integration data computed on board of the FPGA the XScale processor was unsuitable because of the lack of interfaces towards the MAV and the difficulty of allotting well-defined time slots for real-time processing. Therefore, a custom 60 mm × 60 mm PCB featuring an AT91SAM7A3 ARM processor and interface logic has been developed. Its primary objective is the extraction of axis of rotation and rotation rate from the raw wide field integration data, calculated on the FPGA and transmitted towards the ARM processor via CAN bus at full frame rate in simplex mode. As shown by Cuntz et al. (2007) during rotations the fly's VS cell network and its lateral axo-axonal gap junction couplings provide a robust way of encoding the axis of rotation. This *zero-crossing* strategy was implemented on the ARM processor. The axon terminal output of each VS cell was calculated as the weighted sum of the incoming dendritic VS cell data according to the matrix and connection diagram given in Fig. 2. As dendritic VS 10 input into the network the simple sum of pure dendritic VS 10 and HSN values was used. Subsequently the axis of rotation is obtained by determining the zero-crossing location of the resulting ten axon terminal VS cell values. For a crude estimate of the rate and direction of rotation the slope at the zero-crossing location is calculated. Both rotation axis and rate are then further transmitted at full frame rate towards the MAVs flight controller via USART. A USB 2.0 link was also implemented for data transmission towards the MAV or an external PC, e.g., for data logging.

Thus, the interface array on board the PCB consists of two CAN bus ports for accepting incoming data from two independent SmartCam modules, a USART port for communication with the MAV and a microUSB port for MAV communication or optional data logging. An MMC/SD card slot is also provided for future on board data logging, e.g., during flight. For power conversion to the SmartCam's
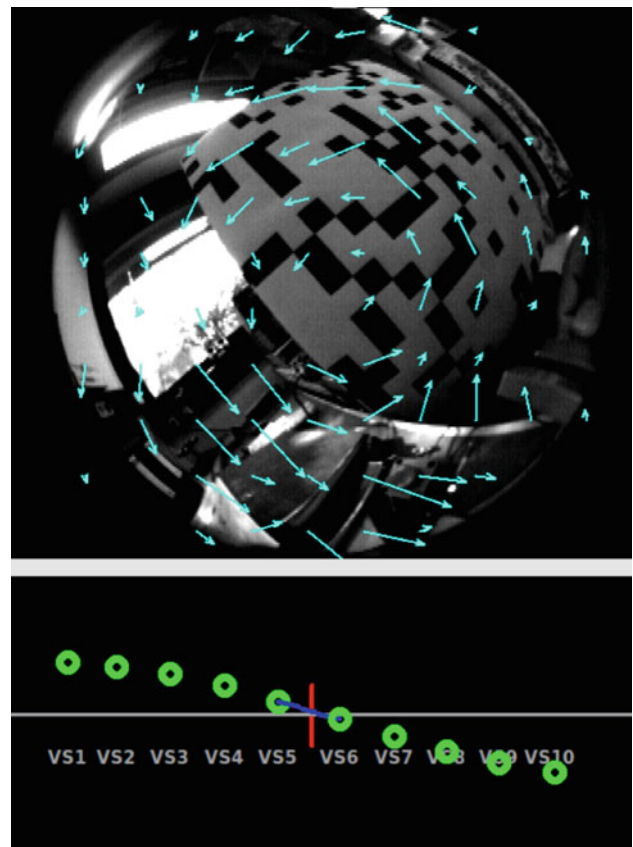

Fig. 4 Screenshot of the GUI on the ground station PC during a rotation around an axis close to zero. The *upper graph* shows the local motion estimates overlaid on top of the live video stream, while the corresponding VS cell network axon terminal responses are displayed in the *lower graph*

voltage requirements of 24 V a Traco Power THB 3-1215 converter (Traco Electronic AG, Zürich, Switzerland) has been included, which additionally strengthens robustness against voltage irregularities owing to motor noise and battery depletion. Total weight of the PCB was 30 g.

3.3.5 PC monitoring software

In order to display and monitor in real time the captured image data along with the estimated local motion and wide field integration data a Linux monitoring interface was written in C++ using the QT framework. For proper visualization in the Graphical User Interface (GUI) the optic flow vectors are scaled and overlaid onto the camera video stream while VS cell homologue data is shown in a corresponding plot (Fig. 4). Data is acquired via an IEEE 802.11g wireless link between the laptop and the sensor system. Due to the limited bandwidth of the wireless connection on average between 10 and 15 frames could be transmitted per second, which nevertheless is sufficient for a human observer to monitor the live image stream and the corresponding optic flow output of

the system. For recording wide field integration data at full frame rate the USB connection between the ARM PCB and the monitoring laptop could be used.

3.4 Design outcome

The final sensor system was able to compute 350 ego-motion estimates per second for transmission towards the MAV flight controller and/or data logging PC, while weighing a total of 193g and consuming less than 4 W off a standard three cell 12 V LiPo RC model battery. At the same time, real time images, flow fields and ego-motion estimates were sent to a control ground station PC at a reduced frame rate of roughly 12 fps. Using automatic gain adaptation the 8 bit image sensor produced pixel values roughly centered in its 0–255 coding range. In dim indoor lighting conditions between 10 and 30 cd/m^2 with an exposure time of 2.85 ms per frame temporal noise caused a typical standard deviation of 2.2% of this range.

4 Results

To test the functionality and reveal the characteristics of the sensor system two kinds of trials were conducted. On one hand, experiments were carried out to ascertain the resemblance with the biological original. On the other hand, essays to elucidate the actual sensor characteristics and accuracy of measurement were performed.

4.1 EMD output characteristics

A distinct feature of correlation type motion detectors is the existence of a velocity optimum in response to moving sine grating stimuli (Buchner 1976; Poggio and Reichardt 1976). As shown by Borst et al. (2003) for a Reichardt Detector

configuration with a temporal high-pass filter in its input lines the velocity response curve for sinusoidal gratings is given by

$$\langle R_i \rangle_\varphi = \Delta I^2 \frac{\tau_{lp} \tau_{hp}^2 \omega^3}{(1 + \tau_{lp}^2 \omega^2)(1 + \tau_{hp}^2 \omega^2)} \sin(2\pi \Delta\varphi/\lambda), \quad (4)$$

where ΔI is the pattern contrast, ω the angular frequency, $\Delta\varphi$ the inter-ommatidial angle, λ the wavelenth and τ_{lp} and τ_{hp} are the low-pass and high-pass time constants, respectively. The velocity optimum is a linear function of the spatial pattern wavelength leading to a constant temporal frequency optimum. This has been observed in behavioral and electrophysiological studies in resting, walking, and flying animals across various fly species. These studies have revealed frequency optima around 1 Hz for stationary *Drosophila*, *Phaenicia*, and *Calliphora* (see Joesch et al. 2008; Eckert 1980; Haag et al. 2004, respectively). For walking *Drosophila* optima from 2 to 3 Hz have been shown (Götz and Wenking 1973; Chiappe et al. 2010). In flying animals, optima have been reported between 3 and 10 Hz for *Drosophila* (Duistermars et al. 2007; Fry et al. 2009), 1 to 10 Hz for *Musca* (Borst and Bahde 1987) and 5 to 7 Hz for *Calliphora* (Hausen and Wehrhahn 1989; Jung et al. 2011). We chose to adjust the filter time constants to values yielding a theoretical frequency optimum similar to *Calliphora* during flight, i.e., at 7.3 Hz ($\tau_{lp} = 45$ ms and $\tau_{hp} = 33$ ms). For confirming the existence and location of the velocity optimum of our sensor system we measured the sensor output for vertically moving sine gratings at spatial wavelengths $\lambda = 12°$, $24°$, and $48°$ at different velocities using a cylinder-shaped LED arena as described by Weber et al. (2010). The normalized mean response of VS 7 cells over $n = 23$ trials revealed velocity optima at 85, 175, and 350°/s for $\lambda = 12°$, $24°$, and $48°$, respectively (Fig. 5a). Dividing the velocity optima by the corresponding spatial wavelength, the frequency optima coincide around 7.3 Hz as predicted by the model calculations in Eq. 4 (Fig. 5b).
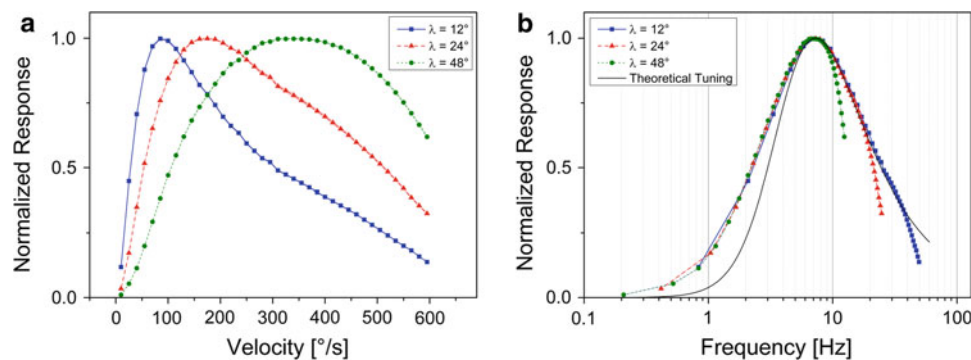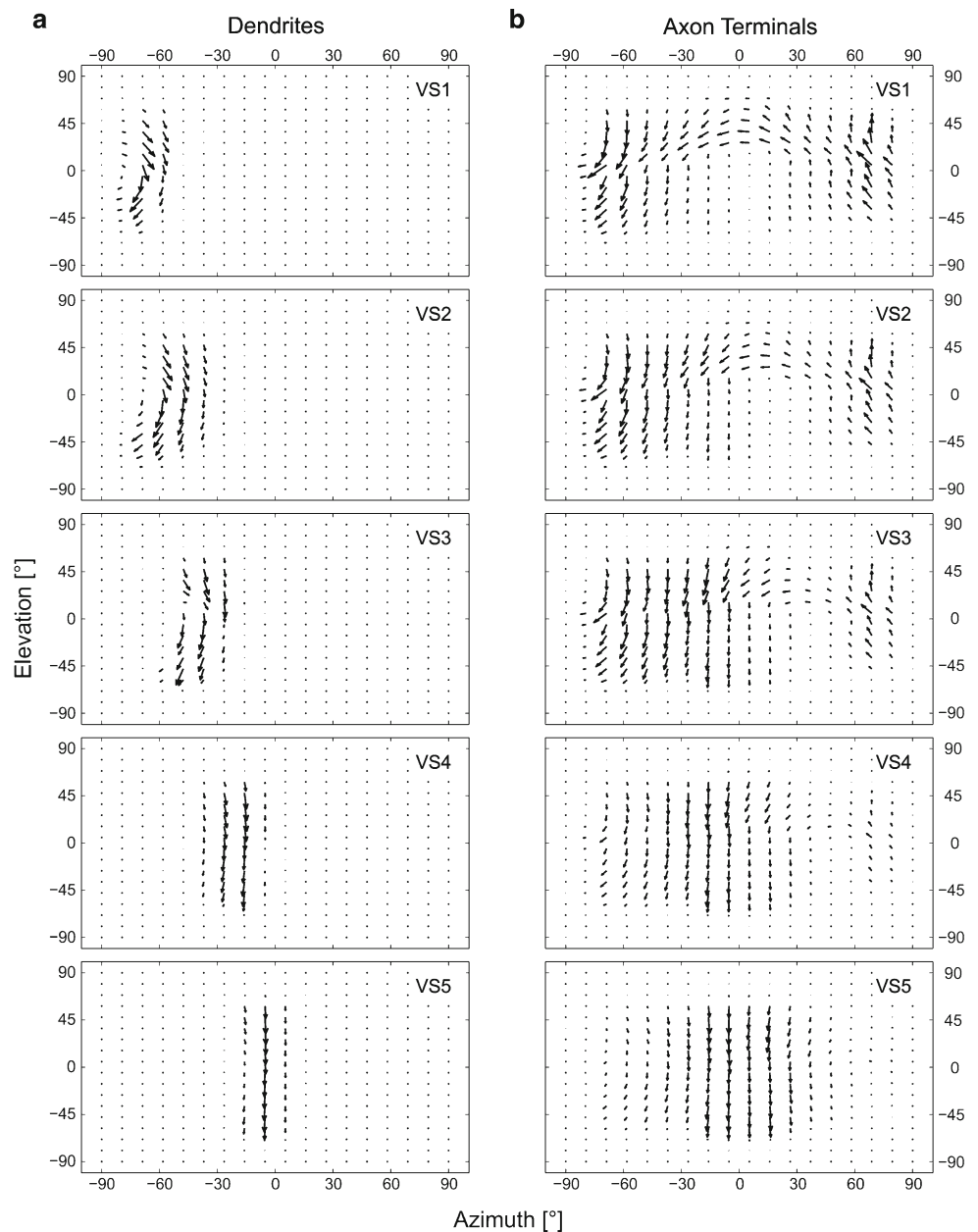

Fig. 5 VS cell velocity tuning. **a** Normalized VS cell response curves for sinusoidal gratings of spatial wavelengths $\lambda = 12°$, $24°$, and $48°$ of visual angle as seen by the fly over a wide range of velocities, displaying optima at 85, 175, and 350°/s, respectively. Representative data from VS 7 cells over $n = 23$ trials are shown. Standard deviation is within symbol size. **b** VS cell frequency tuning. Data from **a** are plotted as a function of temporal frequency by dividing velocity by the respective pattern wavelength. Frequency optima for all three wavelengths lie around 7.3 Hz. Standard deviation is within symbol size
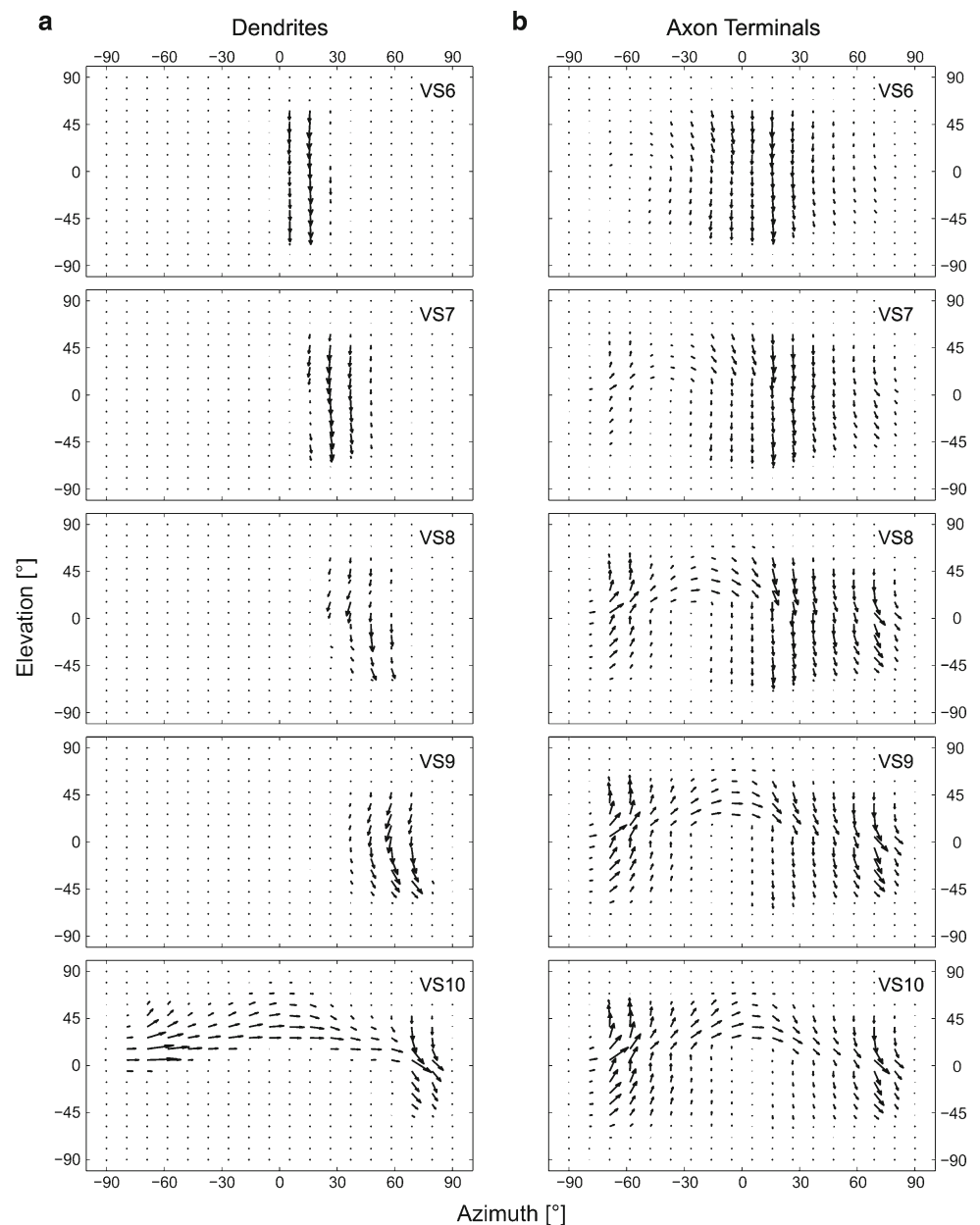
4.2 Receptive fields

To measure our sensor's receptive fields, we used a custom built arena consisting of three 120 Hz SyncMaster 2233RZ monitors (Samsung, South Korea) in combination with NVIDIA GeForce GTX 480/580 graphics cards (NVIDIA Corp., California, USA) using NVIDIA 3D Vision Surround technology. The monitors were perpendicularly placed in front, left, and right of the sensor, covering the visual space from $-135°$ to $135°$ in equatorial azimuth and from $-58°$ to $58°$ in frontal elevation. Monitor radiance inhomogeneities were largely compensated by using diffusing filter paper. Stimuli were written in Python and rendered using the open

source graphics engine Panda3D (Goslin and Mine, 2004). They consisted of a small $5.7° \times 5.7°$ white square moving at $120°/s$ on a dark background (92% Michelson contrast) sweeping the extent of the arena in the frontal hemisphere up and down to reveal the vertical components and left and right to reveal the horizontal components of the receptive fields. For $n = 21$ complete trials the response data was spatially divided into 36×36 equisolid angle bins covering the frontal hemisphere. Both dendritic and axon terminal responses from all implemented VS and HS cells were recorded at the camera frame rate of 350 measurements per second. As expected, the dendritic compartments respond to vertical motion in rather narrow stretches of the visual field (Figs. 6a, 7a), while in

Fig. 7 Receptive fields of VS 6 to VS 10 for **a** dendritic and **b** axon terminal regions. Average data from $n = 21$ trials are shown. Note the resemblance between axon terminal receptive fields and rotations around axes sequentially arranged on the equator

the axon terminals the cells exhibit responses over a much broader area (Figs. 6b, 7b) owing to the lateral connections with their neighboring cells, as depicted in Fig. 2. Thus, the axon terminal output receptive fields resembled matched filters for optic flow generated by rotations around axes sequentially arranged along the equator (see Krapp et al. 1998; Franz and Krapp 2000).

4.3 Rotation axis estimation

The main objective of this sensor system is the estimation of the axis of rotation during ego-motion. In order to examine the measurement accuracy of the sensor system we tested it both in a simulation environment and in a real world scenario.

The experimental setup used for the simulation environment was the same as for the receptive field analysis described in Sect. 4.2. For spatially correct stimulus presentation *cube mapping* was used (Greene 1986). The sensor system was mounted in the focal spot and the simulated environment was rotated around axes ranging from $\theta_{\text{ref}} = -60°$ to $60°$ along the equator in $15°$ steps at angular velocities from 30 to $100°/s$ in $5°/s$ steps for each axis. For performance evaluation the Root Mean Square Deviation (RMSD) between sensor axis estimate and reference angle was defined as

$$\text{RMSD}(\theta_{\text{est}}, \theta_{\text{ref}}) = \sqrt{\frac{\sum_{i=1}^{n} (\theta_{\text{est}}(i) - \theta_{\text{ref}}(i))^2}{n}}, \qquad (5)$$
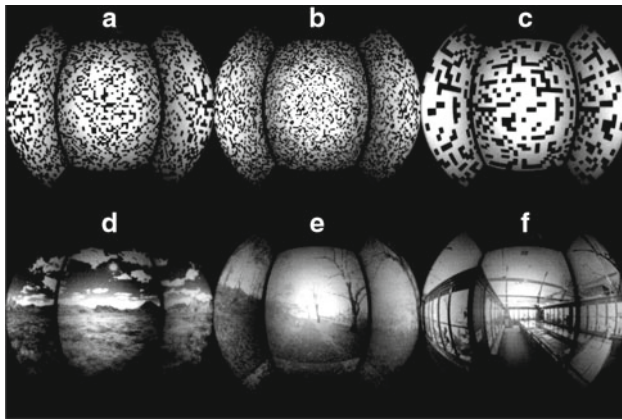
Fig. 8 Screenshots of visual scenarios used in the simulation environment for rotation estimation experiments. Scenes A, B, and C were artificially generated using random probability distributions while scenes D, E, and F were cube mapping projections of photographic scenes

Table 1 Contrast and RMSD for simulated scenes

Scene	Type	Contrast			Result
		C_{RMS}	C_{MAD}	C_{RAPS}	RMSD
A	Artificial	1.39	0.28	0.19	3.00°
B	Artificial	1.36	0.26	0.17	3.65°
C	Artificial	1.38	0.31	0.16	4.59°
D	Natural	1.70	0.17	0.06	9.01°
E	Natural	1.39	0.20	0.04	14.44°
F	Natural	1.42	0.20	0.07	18.42°

where θ_{est} is the sensor axis estimate and θ_{ref} the reference angle.

Three contrast metrics were defined for the presented images. RMS contrast (C_{RMS}) was defined as the standard deviation of pixel values divided by their mean. As a second metric MAD contrast (C_{MAD}) was defined as the Mean Absolute Deviation (MAD) of pixel values. These two metrics are global measures and therefore do not depend on the spatial frequency content or the spatial brightness distribution. Hence, radially averaged power spectrum contrast (C_{RAPS}) was defined as the square root of the mean of the radially averaged power spectra between 0.0649 cycles per degree and 0.6486 cycles per degree, thereby covering the spatial coding range of the sensor system up to its Nyquist limit.

Three artificial and three naturalistic scenarios were presented in the simulation environment (Fig. 8). Owing to their high contrast ratios the artificial scenarios yielded very robust and exact estimation of the rotational axes, on average deviating by less than 5° from the actual axis of rotation (see Table 1). Natural images displayed larger RMS deviations between 9° and 18° due to their lower contrast and the relatively low mean luminance values of available scenes in the experimental setup. In line with the motion detection model, the higher
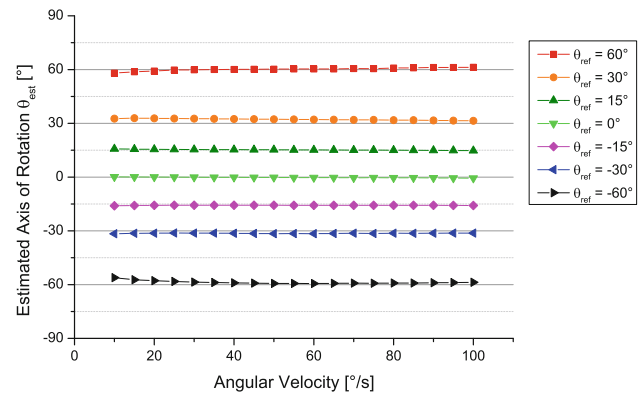

Fig. 9 Axis estimation during real world trials. The sensor system was rotated in a wide range of angular velocities around different reference axes, yielding highly accurate rotation axis estimates virtually regardless of angular velocity

the contrast the better the sensor system performed. To test an extreme case and an exception to this rule we included a scene with high rotational asymmetries (Table 1, row F). In this case the sensor generated the highest RMS deviations in the test set despite not having the lowest contrast. Remarkably, this worst case RMSD of 18.42° is almost identical to the 18.5° spacing between VS cell dendritic receptive fields. This means that the worst case sensor inaccuracy tends to be at most one cell to the left or to the right of the true center of rotation.

For recording data in a real world environment the sensor system was mounted axially on a PLE40 planetary gear with a 100:1 gear reduction ratio (Neugart, Kippenheim, Germany) and rotated using a PANdrive PD1-140-42-SE-232 motor with an integrated control unit (Trinamic, Hamburg, Germany). Situated in an indoor environment with both high and low visual contrast areas the sensor system was rotated around the axes $\theta_{ref} = -60°, -30°, -15°, 0°, 15°, 30°$, and $60°$ along the equator at angular velocities ranging from 10 to 100°/s in 5°/s steps. The actual angular velocity was monitored utilizing the integrated encoders. As can be observed in Fig. 9 rotation axis estimation by the sensor system accurately reflects the actual axis of rotation, basically regardless of rotational velocity. There tends to be, however, slightly higher accuracy towards higher velocities, as can also be seen in Fig. 9.

4.4 Rotation rate estimation

Concurrently with the rotation axis estimation, the sensor system also computes an estimate of the rate of rotation around that particular axis by calculating the slope at the zero-crossing location. To analyze the properties of these estimates, the camera system was subjected to $n = 22$ trials with rotations around the rostro-caudal axis at angular velocities ranging from 10 to 1,000°/s using the same scenes
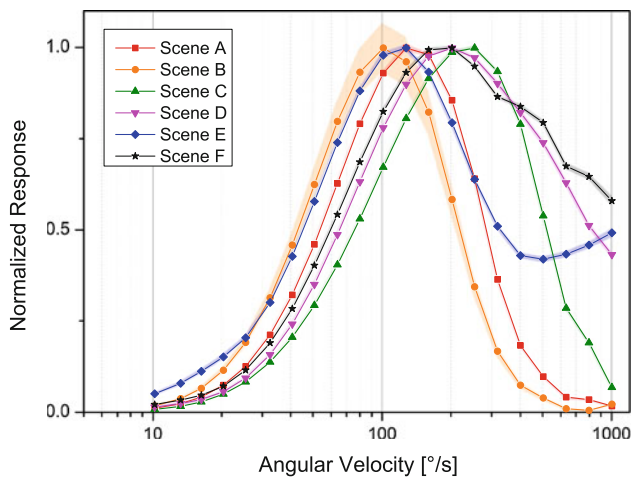
Fig. 10 Rotation rate estimation for naturalistic and artificial scenes. *Traces* show mean estimates while *shaded areas* indicate standard deviation between $n = 22$ trials. Velocity optima lie between 100 and 251°/s

as for the axis estimation trials in Sect. 4.3. Figure 10 presents the typical bell-shaped response curves of the system. Peak responses were found between 100 and 251°/s.

5 Discussion and conclusions

We have presented a fly-inspired visual rotation sensor capable of accurate measurements in a variety of visual scenes, while maintaining the tight restrictions of space, weight and power requirements necessary for use on board MAVs. The outcome of our experiments substantiates on one hand the close resemblance of our implementation with the biological original. On the other hand, our results demonstrate the good performance over a wide range of different visual environments. Our implementation was specifically designed to be used on-board MAVs and therefore features only a small footprint in terms of size, weight and power consumption while maintaining mechanical robustness.

One possible source of performance degradation is the inherent barrel distortion of fisheye lenses. We have therefore tested various correction algorithms on board the FPGA with slight, albeit not substantial performance improvements. This suggests that fisheye lens distortion does not decisively perturb sensor performance.

A particularly useful property of our FPGA- and ARM-based implementation is the versatility and ease of adding other functionalities. By adding different templates for global motion integration, new uses of this sensor system could arise. The simple sum of all VS cell templates for instance, could be used as a measure of global vertical motion for lift control. The sum of HS cells might be used for indication of global horizontal motion for yaw control. Using specialized horizontal templates, *tunnel centering* behavior (Srinivasan et al. 1999) could be implemented for autonomous robot navigation. Along these same lines, collision detection can

be envisioned by using templates for radial expansion. The advantage of our system lies in the fact that there are sufficient free resources for all these computations to be implemented simultaneously, at full frame rate and resolution.

The system has been designed to be used with either one or two cameras, potentially covering the complete 4π sr unit sphere of visual space for true global motion integration and consequently added robustness of ego-motion estimates. Also multi-modal integration of other sensors, such as rate gyroscopes, accelerometers, etc., is supported. This is particularly useful for future studies on sensor fusion with inertial data, akin to integration of vision and inertial haltere measurements in the fly brain. The system we have presented here might therefore prove useful when employed complementarily to inertial measurement units.

In conclusion, we have shown a successful implementation of visual ego-rotation sensing based on the fly visual system, while keeping within tight space, weight and performance restriction boundaries.

References

Aubepart F, El Farji M, Franceschini N (2004) FPGA implementation of elementary motion detectors for the visual guidance of micro-air-vehicles. In: Proc IEEE Int Symp Ind Electron, vol 1, pp 71–76

Autrum H (1952) Über zeitliches Auflösungsvermögen und Primärvorgänge im Insektenauge. Naturwissenschaften 39:290–297

Bermudezi Badia S, Pyk P, Verschure PFM (2007) A fly-locust based neuronal control system applied to an unmanned aerial vehicle: the invertebrate neuronal principles for course stabilization, altitude control and collision avoidance. Int J Robot Res 26:759–772

Beyeler A, Zufferey J, Floreano D (2009) Vision-based control of near-obstacle flight. Auton Robot 27:201–219

Borst A (2007) Correlation versus gradient type motion detectors: the pros and cons. Philos Trans R Soc Lond B 362:369–374

Borst A, Bahde S (1987) Comparison between the movement detection systems underlying the optomotor and the landing response in the housefly. Biol Cybern 56:217–224

Borst A, Reisenman C, Haag J (2003) Adaptation of response transients in fly motion vision. II: Model studies. Vis Res 43:1311–1324

Borst A, Haag J, Reiff DF (2010) Fly motion vision. Annu Rev Neurosci 33:49–70

Buchner E (1976) Elementary movement detectors in an insect visual system. Biol Cybern 24:85–101

Chiappe ME, Seelig JD, Reiser MB, Jayaraman V (2010) Walking modulates speed sensitivity in *Drosophila* motion vision. Curr Biol 20:1470–1475

Conroy J, Gremillion G, Ranganathan B, Humbert JS (2009) Implementation of wide-field integration of optic flow for autonomous quadrotor navigation. Auton Robot 27:189–198

Cuntz H, Haag J, Forstner F, Segev I, Borst A (2007) Robust coding of flow-field parameters by axo-axonal gap junctions between fly visual interneurons. Proc Natl Acad Sci USA 104:10229–10233

Duistermars BJ, Chow DM, Condro M, Frye MA (2007) The spatial, temporal and contrast properties of expansion and rotation flight optomotor responses in *Drosophila*. J Exp Biol 210:3218–3227

Eckert H (1980) Functional properties of the H1-neurone in the third optic ganglion of the blowfly, *Phaenicia*. J Comp Physiol A 135:29–39

Egelhaaf M, Borst A, Reichardt W (1989) Computational structure of a biological motion-detection system as revealed by local detector analysis in the fly's nervous system. J Opt Soc Am A 6:1070–1087

Elyada YM, Haag J, Borst A (2009) Different receptive fields in axons and dendrites underlie robust coding in motion-sensitive neurons. Nat Neurosci 12:327–332

Farrow K, Borst A, Haag J (2005) Sharing receptive fields with your neighbors: tuning the vertical system cells to wide field motion. J Neurosci 25:3985–3993

Franceschini N, Pichon JM, Blanes C, Brady JM (1992) From insect vision to robot vision. Philos Trans R Soc Lond B 337:283–294

Franz MO, Krapp HG (2000) Wide-field, motion-sensitive neurons and matched filters for optic flow fields. Biol Cybern 83:185–197

Fry SN, Rohrseitz N, Straw AD, Dickinson MH (2009) Visual control of flight speed in *Drosophila melanogaster*. J Exp Biol 212:1120–1130

Gonzalez RC, Woods RE (2007) Digital image processing, 3rd edn. Prentice Hall, Upper Saddle River, NJ

Goslin M, Mine MR (2004) The Panda3D graphics engine. Computer 37:112–114

Götz KG, Wenking H (1973) Visual control of locomotion in the walking fruitfly *Drosophila*. J Comp Physiol A 85:235–266

Greene N (1986) Environment mapping and other applications of world projections. IEEE Comput Graph Appl 6:21–29

Haag J, Borst A (2003) Orientation tuning of motion-sensitive neurons shaped by vertical–horizontal network interactions. J Comp Physiol A 189:363–370

Haag J, Borst A (2004) Neural mechanism underlying complex receptive field properties of motion-sensitive interneurons. Nat Neurosci 7:628–634

Haag J, Borst A (2005) Dye-coupling visualizes networks of large-field motion-sensitive neurons in the fly. J Comp Physiol A 191:445–454

Haag J, Borst A (2007) Reciprocal inhibitory connections within a neural network for rotational optic-flow processing. Front Neurosci 1:111–121

Haag J, Denk W, Borst A (2004) Fly motion vision is based on reichardt detectors regardless of the signal-to-noise ratio. Proc Natl Acad Sci USA 101:16333–16338

Harrison RR (2005) A biologically inspired analog IC for visual collision detection. IEEE Trans Circuits Syst I 52:2308–2318

Hassenstein B, Reichardt W (1956) Systemtheoretische Analyse der Zeit-, Reihenfolgen- und Vorzeichenauswertung bei der Bewegungsperzeption des Rüsselkäfers *Chlorophanus*. Z Naturforsch B 11:513–524

Hausen K, Wehrhahn C (1989) Neural circuits mediating visual flight control in flies. I. Quantitative comparison of neural and behavioral response characteristics. J Neurosci 9:3828–3836

Hyslop A, Krapp HG, Humbert JS (2010) Control theoretic interpretation of directional motion preferences in optic flow processing interneurons. Biol Cybern 103:353–364

Joesch M, Plett J, Borst A, Reiff DF (2008) Response properties of motion-sensitive visual interneurons in the lobula plate of *Drosophila melanogaster*. Curr Biol 18:368–374

Jung SN, Borst A, Haag J (2011) Flight activity alters velocity tuning of fly motion-sensitive neurons. J Neurosci 31:9231–9237

Kendoul F, Fantoni I, Nonami K (2009) Optic flow-based vision system for autonomous 3D localization and control of small aerial vehicles. Robot Auton Syst 57:591–602

Köhler T, Röchter F, Lindemann JP, Möller R (2009) Bio-inspired motion detection in an FPGA-based smart camera module. Bioinspir Biomim 4:015,008 (10 pp)

Krapp HG, Hengstenberg B, Hengstenberg R (1998) Dendritic structure and receptive-field organization of optic flow processing interneurons in the fly. J Neurophysiol 79:1902–1917

Land MF, Eckert H (1985) Maps of the acute zones of fly eyes. J Comp Physiol A 156:525–538

Moeckel R, Liu S (2009) Motion detection chips for robotic platforms. In: Floreano D, Zufferey J, Srinivasan MV, Ellington C (eds) Flying insects and robots. Springer, Berlin, pp 101–114

Netter T, Franceschini N (2002) A robotic aircraft that follows terrain using a neuromorphic eye. In: Proceedings of the International Conference on Intellectual Robotic System, vol 1, pp 129–134

O'Carroll DC, Shoemaker PA, Brinkworth RSA (2006) Bio-inspired optical rotation sensor. In: Proceedings of the SPIE Symposium on Smart Structures, Devices, and Systems III, vol 641418, pp 1–12

Petrowitz R, Dahmen H, Egelhaaf M, Krapp HG (2000) Arrangement of optical axes and spatial resolution in the compound eye of the female blowfly *Calliphora*. J Comp Physiol A 186:737–746

Poggio T, Reichardt W (1976) Visual control of orientation behaviour in the fly. Part II. Towards the underlying neural interactions. Q Rev Biophys 9:377–438

Potters M, Bialek W (1994) Statistical mechanics and visual signal processing. J Phys I 4:1755–1775

Reichardt W, Egelhaaf M (1988) Properties of individual movement detectors as derived from behavioural experiments on the visual system of the fly. Biol Cybern 58:287–294

Seitz G (1968) Der Strahlengang im Appositionsauge von *Calliphora erythrocephala* (Meig.). J Comp Physiol A 59(2):205–231

Single S, Borst A (1998) Dendritic integration and its role in computing image velocity. Science 281:1848–1850

Srinivasan MV, Chahl JS, Weber K, Venkatesh S, Nagle MG, Zhang SW (1999) Robot navigation inspired by principles of insect vision. Robot Auton Syst 26:203–216

Srinivasan M, Thurrowgood S, Soccol D (2009) Competent vision and navigation systems. IEEE Robot Autom Mag 16:59–71

Stumpf P (1911) Über die Abhängigkeit der visuellen Bewegungsrichtung und negativen Nachbildes von den Reizvorgängen auf der Netzhaut. Z Psychol 59:321–330

Tatler B, O'Carroll DC, Laughlin SB (2000) Temperature and the temporal resolving power of fly photoreceptors. J Comp Physiol A 186:399–407

Valette F, Ruffier F, Viollet S, Seidl T (2010) Biomimetic optic flow sensing applied to a lunar landing scenario. In: Proceedings of the IEEE International Conference on Robotics and Automation, pp 2253–2260

Weber F, Machens CK, Borst A (2010) Spatiotemporal response properties of optic-flow processing neurons. Neuron 67:629–642

Wertz A, Gaub B, Plett J, Haag J, Borst A (2009) Robust coding of ego-motion in descending neurons of the fly. J Neurosci 29:14993–15000

XILINX (1999) XAPP134: Virtex synthesizable high performance SDRAM controller. XILINX

XILINX (2003) XAPP394: Interfacing to mobile SDRAM with CoolRunner-II CPLDs. XILINX

Zhang T, Wu H, Borst A, Kühnlenz K, Buss M (2008) An FPGA implementation of insect-inspired motion detector for high-speed vision systems. In: Proceedings of the IEEE International Conference on Robotics and Automation, pp 335–340

Zufferey J, Floreano D (2006) Fly-inspired visual steering of an ultra-light indoor aircraft. IEEE Trans Robot 22:137–146