Biol Cybern
DOI 10.1007/s00422-006-0097-1

ORIGINAL PAPER

Biological Cybernetics

Depth estimation using the compound eye of dipteran flies

Konstantinos Bitsakos · Cornelia Fermüller

Received: 28 November 2005 / Accepted: 5 July 2006

Abstract In the neural superposition eye of a dipteran fly every ommatidium has eight photoreceptors, each associated with a rhabdomere, two central and six peripheral, which altogether result in seven functional light guides. Groups of eight rhabdomeres in neighboring ommatidia have largely overlapping fields of view. Based on the hypothesis that the light signals collected by these rhabdomeres can be used individually, we investigated the feasibility of estimating 3D scene information. According to Pick (Biol Cybern 26:215–224, 1977) the visual axes of these rhabdomeres are not parallel, but converge to a point 3–6 mm in front of the cornea. Such a structure theoretically could estimate depth in a very simple way by assuming that locally the image intensity is well approximated by a linear function of the spatial coordinates. Using the measurements of Pick (Biol Cybern 26:215–224, 1977) we performed simulation experiments to find whether this is practically possible. Our results indicate that depth estimation at small distances (up to about 1.5–2 cm) is reasonably accurate. This would allow the insect to obtain at least an ordinal spatial layout of its operational space when walking.

1 Introduction

In the biological world a large variety of eye designs exists. It has been estimated that eyes have evolved no fewer than 40 times, independently, in different parts of the animal kingdom (Dawkins 1996). An adaptationist's logic leads to the conclusion that these eye designs and therefore the images they capture, are highly adapted to the tasks the animal has to perform. For many years neurobiologists have performed comparative studies of eyes at the physical level. Some examples are investigations of the physical limits of visual discrimination (Barlow 1964) and comparisons of factors such as resolving power and sensitivity (Land 1981). It is well understood that eyes always embody in their design complicated compromises between different competing physical factors. Recently, in the literature of Robotics and Computer Vision, an effort started to design new camera systems which are better suited than standard cameras for solving specific tasks. Along with these efforts, computational researchers also started to compare the performance of different eyes or cameras at a higher level, that is at the level of the computational tasks. To give an example, it was shown that eyes covering a full sphere have a computational advantage over eyes with smaller field of view for the task of visual 3D motion interpretation (Dahmen et al. 1997; Fermüller and Aloimonos 2000).

Here, we take a look at the compound eye of flies. Is there an advantage to the particular arrangement of many ommatidia, each with a few photoreceptors and each with its own lens? Or is this implementation simply a cost-effective solution? We investigate whether this architecture allows for easy distance estimation. Ordinary stereo between the two eyes, as it has been found in the praying mantis (Rossel 1983) does not seem likely, because there is hardly any overlap of the images in the outward pointing compound eyes of flies. Regular stereo between neighboring ommatidia of individual eyes,

K. Bitsakos (✉) · C. Fermüller
Computer Vision Laboratory, UMIACS,
University of Maryland, College Park, MD, USA
e-mail: kbits@cs.umd.edu

C. Fermüller
e-mail: fer@cfar.umd.edu

because of the small distances between them, is not meaningful either. Another form of deriving distance may still be possible.

In the compound eye of the dipteran fly, the rhabdomeres of an ommatidium are not fused, as in many other compound eyes, but sample different parts of the image produced by the corneal lens. The eight rhabdomeres, two narrow ones in the center stacked one on top of the other, and six wider ones arranged around the central pair, form a pattern of one central plus six peripheral lightguides (Franceschini 1975). Groups of seven rhabdomeres (counting the central pair as one) each from a different ommatidium possess largely overlapping visual fields. The two central photoreceptors project directly to the medulla, without synapsing in the lamina. The six peripheral receptors send their input to a common cartridge (neurommatidium) in the lamina where the signals are among other things superpositioned to obtain enhanced light sensitivity; thus the name, "neural superposition". It is generally considered that the visual axes of the six rhabdomeres of one neurommatidium are parallel, such that perfect signal enhancement for all spatial wavelengths can occur. Pick [1977], however, found that there is no parallelism. Instead the six visual axes of one neurommatidium converge to a point 3–6 mm in front of the corneal surface. He also showed that the non-parallelism does not conflict with the basic concept of neural superposition. He argued that this special arrangement of the rhabdomeres is an adaptation for high absolute light sensitivity (increasing light sensitivity and reducing optical crosstalk between neighboring rhabdomeres). He also suggested that it may possibly aid in estimating distances of visual objects. Following this thought, in this paper we present a mathematical theory and experiments on how depth could be obtained with such an eye architecture.

We compare the optical system of the fly's eyes to a system of cameras. It is like an array of closely spaced simple cameras with seven pixels each arranged on a sphere (or two hemispheres). In the abstract, camera systems can be studied through the concept of the plenoptic function, which is the function of all the light rays covering a certain space. Any camera system samples a subset of the plenoptic function. In Neumann et al. [2004] it has been shown that theoretically a spherical arrangement of very closely spaced pinhole cameras could allow for easy estimation of distance.

The underlying idea is as follows. Assuming the plenoptic function to be continuous, it can be linearized in a small spatial neighborhood (as we usually do in image motion estimation). We then can compare the change of intensity of adjacent rays through the same focal point with the change of intensity of parallel rays through adjacent focal points. From the comparison of similar triangles, we obtain that their ratio provides distance.

According to the Pick model we have rays from close by scene points forming images on multiple (seven) ommatidia, with all the rays carrying different image information. Thus, using the linearization we could obtain distance. The question is whether the few rays and the low acuity of the fly's eye would actually allow for any reliable estimate. We find it is possible at close distances. In the remainder of the paper we recap the theory on how the distance parameters can be estimated and present results from experiments using synthetic images as those seen by compound eyes.

The paper is organized as follows. Section 2 describes the model of the eye of dipteran flies that we used for our simulations. Section 3 synopsizes the theoretical background and the tools used to analyze the images taken from such an eye. Then a detailed explanation on how depth estimation using these images can be performed, is provided. Section 4 presents a plausible implementation of the estimation algorithm using a neural network and discusses some biological issues. Section 5 describes experiments and an analysis relating the distance estimation to the filtered signal. Finally, Sect. 6 provides a summary and discussions.

2 The model

The compound eye of a dipteran fly (Fig. 1) consists of many single eyes, called *ommatidia*, arranged on the surface of a spheroid. Each ommatidium is adjacent to six other ommatidia and contains eight photoreceptors R_1–R_8 each with a rhabdomere (that contains the visual pigments). In Fig. 2a only seven photoreceptors are de-

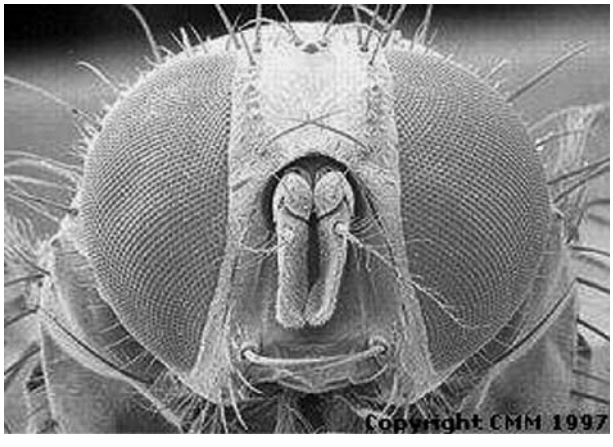


Fig. 1 The compound eye of a housefly (*Musca domestica*) consists of approximately 3,000 ommatidia arranged on a spherical structure. Courtesy of Duncan Waddell

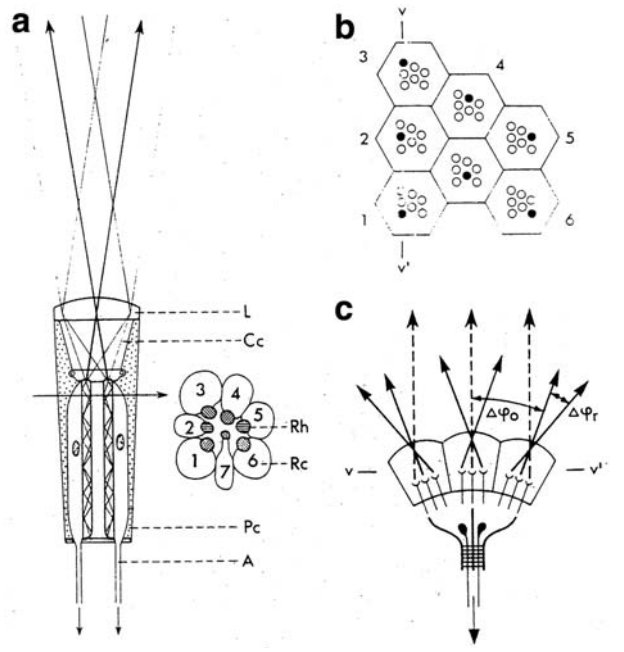


Fig. 2 **a** Schematic longitudinal and transverse section of a fly ommatidium. *A* retinula-cell axon; *Cc* crystalline cone; *L* cornea lens; *Pc* pigment cell; *Rc* retinula cell; *Rh* rhabdomere. **b** *Filled circles* indicate the pattern of retinula cells whose visual field axes are parallel. Dorsal left eye: ommatidia 1, 2, 3 are located medial, 5 and 6 caudal as seen from outside the eye. **c** Illustration of retina-to-lamina projection with one synaptic cartridge. $\Delta\phi_r$ denotes the divergence angle between the visual axes of neighboring rhabdomeres, $\Delta\phi_o$ the interommatidial angle (redrawn from Pick 1977)

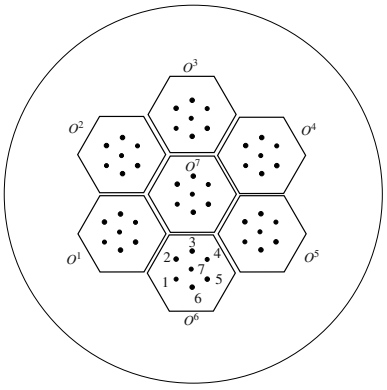


Fig. 3 We model the compound eye of the fly as a collection of planes tangent to the surface of a unit sphere. Each ommatidium is a hexagonal patch and consists of seven photoreceptors (represented by *dots*). In our model, six of them are placed symmetrically around the center of the *hexagon*

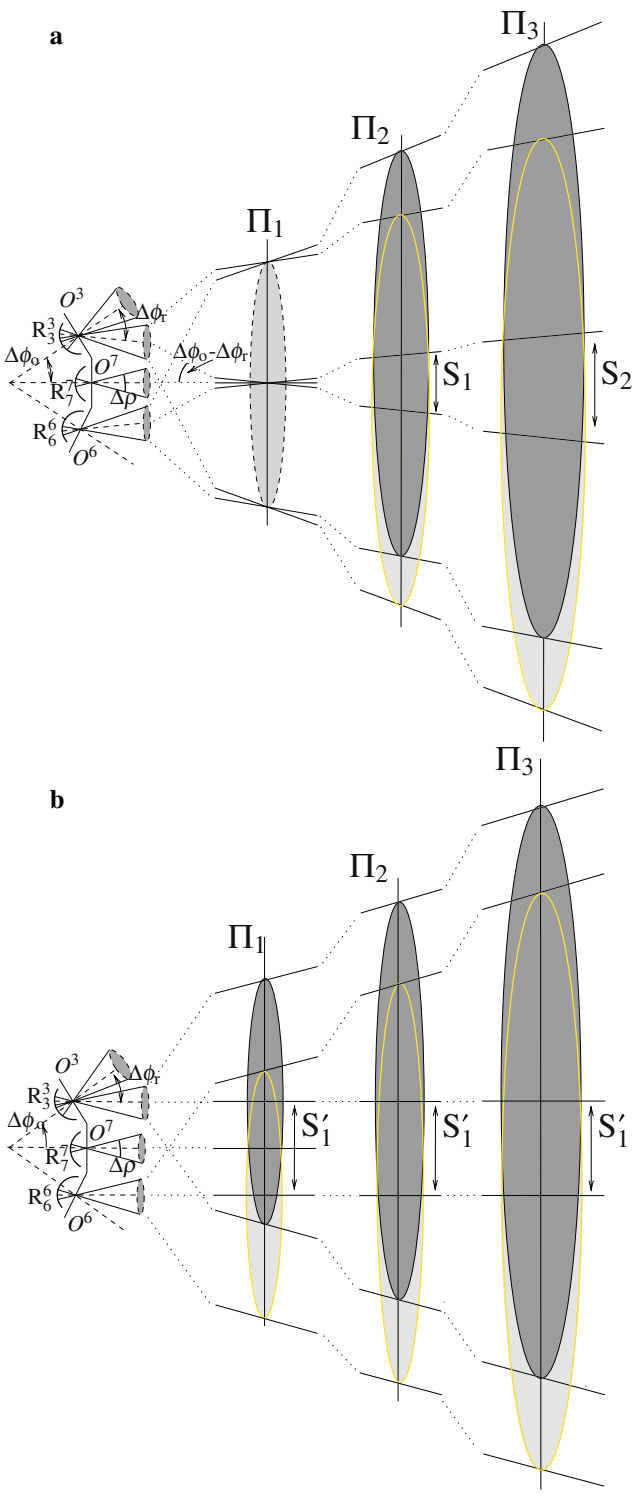
picted, because the rhabdomere of R_8 lies on top of the one of R_7 and has the same visual axis.

The angle formed by the optical axes of the central rhabdomeres of two adjacent ommatidia is denoted by $\Delta\phi_o$, and according to (Pick 1977) is approximately 1.88°.[1] The optical axes of the rhabdomeres within the same ommatidium have a divergence angle $\Delta\phi_r$, which is systematically larger than $\Delta\phi_o$, with values ranging from 2.18° to 2.23° (Pick 1977). As a consequence, the optical axes of the central rhabdomere of an ommatidium and a properly chosen peripheral rhabdomere of an adjacent ommatidium form an angle $\Delta\phi_s = \Delta\phi_r - \Delta\phi_o$, which is about 16 – 19% of $\Delta\phi_o$ (see Figs. 2c, 4a).

For our model it is sufficient to consider a small part of the compound eye consisting of seven ommatidia, a central surrounded by six peripheral ones as shown in Fig. 3. We model each ommatidium as a hexagonal planar patch, tangent to a sphere of unit radius. Photoreceptors R_7 and R_8 are treated as a single one that we denote R_7. For the sake of simplicity, we place the six peripheral photoreceptors (R_1–R_6) symmetrically around the center of the hexagon where R_7 (and R_8) lie. The distance between R_7 and any peripheral photoreceptor is one-half the height of the equilateral triangle (the hexagon is made up of six such triangles). Figure 4 shows a cross-section through the model. Figure 4a sketches the visual fields as seen by the peripheral photoreceptors according to the Pick model, and Figure 4b shows what the visual fields would look like if the axes where parallel.

It is well known that in the compound eye of flies the peripheral photoreceptors (R_1–R_6) and the central photoreceptors (R_7, R_8) exhibit different functional characteristics; achromatic vs chromatic input; better versus worse photon absorbance capability; higher versus lower signal to noise ratio; faster versus slower response and lower versus higher absolute gain, respectively (Anderson and Laughlin 2000). Also, the angular sensitivity and the acceptance (or opening) angle $\Delta\rho$ of the two groups of photoreceptors is different, and it changes with the light level. In our model we do not consider the dependence of the opening angle to the rhabdomere's type, light's wavelength and intensity. We consider $\Delta\rho$ to be a constant approximately equal to the interommatidial angle ($\Delta\rho = 1.9°$). Since we need to combine intensity estimates within the two groups, we assume that there is a normalization step, as has been suggested in Laughlin [1981], that ensures that the ratio of outputs from different classes of photoreceptors remains invariant as a function of the changing light intensity. In Sect. 4 we discuss how the angular sensitivity of the different photoreceptors in the compound eye might affect the performance of the described algorithm.

[1] All the measurements mentioned in Pick [1977] refer to females houseflies (*Musca domestica*).



◀ **Fig. 4** **a** Converging case: The optical axes of the central photoreceptors of adjacent ommatidia diverge at an angle $\Delta\phi_o$, while the optical axes of a central and a peripheral photoreceptor within the same ommatidium diverge at an angle $\Delta\phi_r$ ($>\Delta\phi_o$). As a result the axes of the central rhabdomere of the central ommatidium (R_7^7) and the peripheral rhabdomere of an adjacent ommatidium R_j^j, $j \in \{1 \ldots 6\}$ form an angle $\Delta\phi_r - \Delta\phi_o > 0$. Each photoreceptor's field of view is represented by a cone with opening angle $\Delta\rho$. The intersection plane (Π_1) is located 3–6 mm in front of the ommatidium. The cones of the converging rhabdomeres coincide at Π_1, and as the depth of the scene increases they diverge (denoted by the distance $S_1 < S_2$). **b** The hypothetical case of $\Delta\phi_o = \Delta\phi_r$: the distance of the visual axes through the peripheral photoreceptors, denoted as S'_1, is the same for all depth values (denoted by the scene planes Π_1, Π_2 and Π_3). As a result, for larger depth values S'_1 is much smaller than the diameter of the visual field and can be ignored

3 Plenoptic video geometry

Most work in computational vision is applicable to pinhole cameras and human vision. The compound eye of dipteran flies, as described in Sect. 2, is quite different from the human eye. To analyze compound eye images, we need to extend some concepts and tools used for analyzing images taken by conventional cameras. In this section we will present these concepts.

The visual space is completely described by all the light rays in it. Any camera can be viewed as a device that samples a subset of the space of light rays. A mathematical description of this space is given by the "plenoptic" function as described by Adelson and Bergen [1991]. For every position in space, every instance of time, and every orientation it records the intensity of the light ray or, more generally, the spectral energy at multiple wavelengths. For our purposes we will only consider it a scalar function, which at a certain time instance maps position and orientation to intensity. A number of studies in computer vision and computer graphics have recently used representations of the plenoptic function for rendering visual information, for example the "light field" (Levoy and Hanrahan 1996) and the "lumigraph" (Gortler et al. 1996). In Neumann and Fermüller [2003] the concept of the "polydioptric" camera was introduced. This camera consists of a surface with an array of pinhole cameras at closely spaced positions and allows for easy depth estimation.

We model the compound eye as a very sparse implementation of the polydioptric camera. We use the same concepts, but the sparseness of the image receptors makes the estimation a bit more complicated. The estimation of depth is performed in two computational steps. The first step estimates through interpolation the intensity difference between (hypothetical) parallel rays, and the intensity difference of converging rays along the gradient direction. This amounts to solving an overdetermined system of linear equations (Sect. 3.3). The second step computes the depth as the ratio of the two intensity differences (Sect. 3.4). Together the two steps provide a mathematically solid way, which uses all the

data available and does not depend on the specific regular geometry of the model. However, the neural structure of the fly may instead compute in simpler ways approximations to this estimate. In Sect. 3.5 we discuss an estimation that uses every available direction in the image independently; that is, there is no interpolation step.

Let the image intensity function of interest be a scalar function $\mathrm{L}(\mathbf{x}, \mathbf{r})$, where $\mathbf{x}$ denotes the position, that is the coordinates on the surface, $\mathbf{r}$ denotes the parameters of the orientation of the ray. Both $\mathbf{x}$ and $\mathbf{r}$ are two-dimensional. However, for ease of notation, we use coordinates on a sphere and express them as three-dimensional vectors. Let then $\mathbf{x}$ be the vector from the coordinate center to a point on the surface, and let $\mathbf{r}$ be an orientation vector of unit length.

3.1 Directional derivative $\nabla_{\boldsymbol{r}}\boldsymbol{L}$

First, let us consider the image formed by a spherical perspective camera. The model consists of a spherical surface Π, the image surface, which is one unit distance away from the 3D world point $\mathbf{O}$, the center or center of projection. Let us attach a coordinate system *uvw* to the center of projection. The image is formed on Π by intersecting the pencil of rays through $\mathbf{O}$ with the imaging surface. In this case we are capturing a subset of the space of light rays, namely those passing through a single point ($\mathbf{O}$). Thus, the image can be considered as a function of varying orientation ($\mathbf{r} = (u, v, w)^{\mathrm{T}}$). The partial derivatives of $\mathrm{L}(\mathbf{x}, \mathbf{r})$ with respect to orientation $(\frac{\partial L}{\partial u}, \frac{\partial L}{\partial v}, \frac{\partial L}{\partial w})^{\mathrm{T}} = \nabla_{\boldsymbol{r}}\boldsymbol{L}$ indicate how the intensity at a given point changes as the direction of the light changes.

3.2 Positional derivative $\nabla_{\boldsymbol{x}}\boldsymbol{L}$

Now, let us consider the theoretical concept of an orthographic camera. The model consists of a plane Π, the image plane and an orientation in the 3D world $\mathbf{r}$. Let us attach a coordinate system *xyz* to the center. The image is formed on Π by all rays with orientation $\mathbf{r}$ crossing this plane. So the image at any time can be viewed as a function of varying position ($\mathbf{x} = (x, y, z)^{\mathrm{T}}$). The partial derivatives of $\mathrm{L}(\mathbf{x}, \mathbf{r})$ with respect to position, $(\frac{\partial L}{\partial x}, \frac{\partial L}{\partial y}, \frac{\partial L}{\partial z})^{\mathrm{T}} = \nabla_{\boldsymbol{x}}\boldsymbol{L}$ indicate how the intensity for a given orientation changes as the position of the light ray changes.

3.3 Estimating the derivatives

Our camera model combines elements of both orthographic and perspective cameras. More precisely if we consider the central photoreceptors only, then we have a spherical perspective camera.[2] Similarly, if we consider each ommatidium in isolation, then we have a planar perspective camera. On the other hand, if we think of the almost converging light rays recorded by the photoreceptors of a single neurommatidium, then both the position and the orientation differ, so we have a mixed camera model. From the given image intensity measurements we have to compute the positional ($\nabla_{\boldsymbol{x}}\boldsymbol{L}$) and directional derivatives ($\nabla_{\boldsymbol{r}}\boldsymbol{L}$) at the central photoreceptor of an ommatidium. Assuming that the image intensity is linear in a local neighborhood, we can then obtain these derivatives by considering their projections on the given receptor measurements. That is, we compute the scalar products of the derivative vectors with the difference vectors of orientation and position. The projection of vector $\mathbf{a}$ on $\mathbf{b}$ is written $\mathbf{a} \cdot \mathbf{b}$, with "$\cdot$" denoting the scalar products, and in matrix notation this is expressed as $\mathbf{a}^{\mathrm{T}}\mathbf{b}$. In the following equations we use the indexing as applied in Fig. 3. We denote the ommatidium that a photoreceptor belongs to with a superscript and the position of the photoreceptor inside the ommatidium with a subscript. The central ommatidium and the central photoreceptor are numbered 7 and the rest are numbered 1–6. [3] A ray at photoreceptor R_i^j has orientation $\mathbf{r}_i^j$ and position $\mathbf{x}_i^j$ and L_i^j is the intensity "seen" by the photoreceptor. We then obtain:

$$\underset{\text{photoreceptors}}{\text{central}} \quad \nabla_{\boldsymbol{r}}\boldsymbol{L}^{\mathrm{T}}(\mathbf{r}_7^i - \mathbf{r}_7^7) = \mathrm{L}_7^i - \mathrm{L}_7^7\,, i \in \{1 \ldots 6\} \tag{1}$$

$$\underset{\text{photoreceptors}}{\text{“neuro -ommatidium”}} \quad \begin{pmatrix} \nabla_{\boldsymbol{x}}\boldsymbol{L} \\ \nabla_{\boldsymbol{r}}\boldsymbol{L} \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} \mathbf{x}_{3+j}^{3+j} - \mathbf{x}_j^j \\ \mathbf{r}_{3+j}^{3+j} - \mathbf{r}_j^j \end{pmatrix} = \mathrm{L}_{3+j}^{3+j} - \mathrm{L}_j^j, \quad \text{for } j \in \{1 \ldots 3\}. \tag{2}$$

In addition,

$$\text{radiance constancy along } \mathbf{r} \qquad \nabla_{\boldsymbol{r}}\boldsymbol{L}^{\mathrm{T}}\mathbf{r}_7^7 = 0 \tag{3}$$

$$\text{radiance constancy along } \mathbf{r} \qquad \nabla_{\boldsymbol{x}}\boldsymbol{L}^{\mathrm{T}}\mathbf{r}_7^7 = 0 \tag{4}$$

Equation (1) expresses the relationship between the central photoreceptors ($\mathrm{R}_7^7, \mathrm{R}_7^i, i \in \{1 \ldots 6\}$) of two adjacent ommatidia (O^7, O^i). Equation (2) expresses the

[2] All the rays pass through the center, while the image is formed on the surface of the sphere.

[3] The numbering convention is the same as the one used in Pick [1977], Fig. 2.

relationship between two peripheral photoreceptors of two peripheral ommatidia with converging visual axes (R_4^4–R_1^1, R_5^5–R_2^2 and R_6^6–R_3^3). While the central photoreceptor of the central ommatidium (R_7^7) also has an almost parallel optical axis with the aforementioned photoreceptors it is not included in (2), because only the six peripheral retinula cells send their input to a common synaptic cartridge in the lamina (Trujillo-Cenoz and Melamed 1966; Braitenberg 1967).

Equations (3) and (4) express the fact that a transparent medium such as air does not change the intensity of the light, thus the radiance along the viewing direction $\mathbf{r}$ is constant and both directional and positional derivatives along $\mathbf{r}$ are zero.

The above Eqs. (1) and (2) form a linear system $\mathbf{A} \cdot \mathbf{X} = \mathbf{B}$. If we assume that the central photoreceptor of the central ommatidium (R_7^7) lies on the z axis i.e., has position and orientation $x_7^7 = r_7^7 = (0,0,1)^{\mathrm{T}}$, then the z-component of the positional and the directional derivative is zero ($\frac{\partial L}{\partial z} = \frac{\partial L}{\partial w} = 0$) and we obtain the following system

$$\begin{bmatrix} 0 & 0 & u_7^1 & v_7^1 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & u_7^6 & v_7^6 \\ x_4^4 - x_1^1 & y_4^4 - y_1^1 & u_4^4 - u_1^1 & v_4^4 - v_1^1 \\ x_5^5 - x_2^2 & y_5^5 - y_2^2 & u_5^5 - u_2^2 & v_5^5 - v_2^2 \\ x_6^6 - x_3^3 & y_6^6 - y_3^3 & u_6^6 - u_3^3 & v_6^6 - v_3^3 \end{bmatrix} \cdot \begin{bmatrix} \frac{\partial L}{\partial x} \\ \frac{\partial L}{\partial y} \\ \frac{\partial L}{\partial u} \\ \frac{\partial L}{\partial v} \end{bmatrix} = \begin{bmatrix} \mathrm{L}_7^1 - \mathrm{L}_7^7 \\ \vdots \\ \mathrm{L}_7^6 - \mathrm{L}_7^7 \\ \mathrm{L}_4^4 - \mathrm{L}_1^1 \\ \mathrm{L}_5^5 - \mathrm{L}_2^2 \\ \mathrm{L}_6^6 - \mathrm{L}_3^3 \end{bmatrix} \tag{5}$$

$\mathbf{X} = (X_j)$ is a 4×1 vector containing the unknown derivatives (two directional and two positional derivatives), $\mathbf{A} = (\mathrm{a}_{i,j})$ is a 9×4 matrix and $\mathbf{B} = (\mathrm{B}_i)$ is a 9×1 vector containing the differences in the intensity values. In the ideal case, this system is clearly over-determined, as only four linear independent equations are needed to solve it. In practice, the intensity values of vector $\mathbf{B}$ are not perfect, so we need to find the "best" estimate for $\mathbf{X}$ by minimizing the error. We use the sum of squared differences as the cost function and solve the least squares minimization

$$\min_{\mathbf{X}\in\mathbb{R}^4} \|\mathbf{AX} - \mathbf{B}\|_2^2 = \min_{\mathbf{X}\in\mathbb{R}^4} \sum_{i=1}^{9} \left(\sum_{j=1}^{4} \mathrm{a}_{ij} x_j - \mathrm{b}_i \right)^2 . \tag{6}$$

The four components of $\mathbf{X}$ are not independent. From Eqs. (3) and (4) we obtain that the directional and the positional derivative vectors are parallel. This relation can be enforced by requiring the cross-product of the two vectors to be zero. The optimization problem then takes the form:

$$\begin{aligned} &\min_{\mathbf{X}\in\mathbb{R}^4} \|\mathbf{AX} - \mathbf{B}\|_2^2 \\ &\text{subject to} \quad \nabla_{\boldsymbol{r}} \boldsymbol{L} \times \nabla_{\boldsymbol{x}} \boldsymbol{L} = \mathbf{0}. \end{aligned} \tag{7}$$

The last constraint makes the optimization problem non-linear and thus more difficult to solve. In our experiments we found that it does not lead to significantly better results. Thus, for a simple implementation, it is not necessary to enforce this constraint.

Before going on, let us consider also how our equations would look like if the visual axes of the photoreceptors of an neurommatidium were parallel. Then in Eq. (2)

$$\mathbf{r}_{3+j}^{3+j} - \mathbf{r}_j^j = 0 \quad \text{for all} \quad j \in \{1 \ldots 3\},$$

and thus Eq. (2) simplifies to

$$\left(\nabla_{\boldsymbol{x}} \boldsymbol{L}\right)^{\mathrm{T}} \left(\mathbf{x}_{3+j}^{3+j} - \mathbf{x}_j^j \right) = \mathrm{L}_{3+j}^{3+j} - \mathrm{L}_j^j, \quad \text{for } j \in \{1 \ldots 3\} \tag{8}$$

In this case Eqs. (1) and (2) are independent. The positional derivatives are simply interpolated from the rays of the peripheral receptors, and the directional derivatives are interpolated from the rays of the central receptors.

3.4 Distance estimation

Using the estimated values for the positional and directional derivatives ($\nabla_{\boldsymbol{x}} \boldsymbol{L}$, $\nabla_{\boldsymbol{r}} \boldsymbol{L}$), one can approximate the distance of the object viewed. [4] This relationship can be obtained from the law of similar triangles. It was first derived in Bolles et al. [1987] for differential stereo and image plane analysis. The derivation is shown next.

Let us attach a coordinate system xyz for position and a coordinate system uvw for orientation to the center of

[4] Let us clarify: the term depth is usually used to refer to the distance along the z-axis in planar cameras. In spherical imaging systems, as in the eye of the fly, we are computing the distance from the center of the sphere to the object, which is also called the range. In this paper we have used loosely the term depth, as locally the distance was measured along the z-axis. Globally, however, with respect to the whole eye, these measurements always refer to distance.

our spherical camera **O**. The two coordinate systems are aligned. Consider a plane $\Pi_{\mathrm{O}} : z = 0$ at the center and a plane $\Pi_{\mathrm{R}} : z = 1$ tangent to the unit sphere. The camera sees a planar object Π_{Obj}, which we assume to be parallel to Π_{O} and Π_{R}, located Z_0 units away from the center **O**. Consider three points P_0, P_1 and P_2 on the object, which are viewed by the rays $\mathrm{L}_{\mathrm{P}_0} = \mathrm{L}(\mathbf{x}, \mathbf{r})$, $\mathrm{L}_{\mathrm{P}_1} = \mathrm{L}(\mathbf{x} + \mathrm{d}\mathbf{x}, \mathbf{r})$ and $\mathrm{L}_{\mathrm{P}_2} = \mathrm{L}(\mathbf{x}, \mathbf{r} + \mathrm{d}\mathbf{r})$ as shown in Fig. 5. That is, the change in orientation between rays through P_0 and P_2 is $\mathrm{d}\mathbf{r} = (\mathrm{d}u, \mathrm{d}v, 0)^{\mathrm{T}}$, and the change in position between rays through P_0 and P_1 is $\mathrm{d}\mathbf{x} = (\mathrm{d}x, \mathrm{d}y, 0)^{\mathrm{T}}$, with $\mathbf{r} \cdot \mathrm{d}\mathbf{x} = 0$ and $\mathbf{r} \cdot \mathrm{d}\mathbf{r} = 0$. Figure 5 shows a cross-section parallel to the y-axis through the setting, thus showing only the x and u components.

Since

$$\mathrm{d}x = \mathrm{d}u \tag{9}$$

we have that

$$\mathrm{d}x = \|\overrightarrow{\mathrm{P}_O\mathrm{P}_1}\| = \frac{1}{Z_0}\|\overrightarrow{\mathrm{P}_0\mathrm{P}_2}\| \quad \text{(similar triangles argument)}.$$

Assuming that the intensity in the object plane varies linearly, we obtain

$$\frac{\mathrm{L}_{\mathrm{P}_1} - \mathrm{L}_{\mathrm{P}_0}}{\mathrm{L}_{\mathrm{P}_2} - \mathrm{L}_{\mathrm{P}_0}} = \frac{\|\overrightarrow{\mathrm{P}_0\mathrm{P}_1}\|}{\|\overrightarrow{\mathrm{P}_0\mathrm{P}_2}\|} \Rightarrow$$

$$\frac{\mathrm{L}(x + \mathrm{d}x, u) - \mathrm{L}(x, u)}{\mathrm{L}(x, u + \mathrm{d}u) - \mathrm{L}(x, u)} = \frac{\|\overrightarrow{\mathrm{P}_0\mathrm{P}_1}\|}{\|\overrightarrow{\mathrm{P}_0\mathrm{P}_2}\|} = \frac{1}{Z_0}.$$

Taking the limit of the above equation as $\mathrm{d}x$ and $\mathrm{d}u$ go to zero we obtain

$$\lim_{\mathrm{d}u, \mathrm{d}x \to 0} \frac{\mathrm{L}(x + \mathrm{d}x, u) - \mathrm{L}(x, u)}{\mathrm{L}(x, u + \mathrm{d}u) - \mathrm{L}(x, u)} = \frac{L_x}{L_u} = \frac{1}{Z_0}.$$

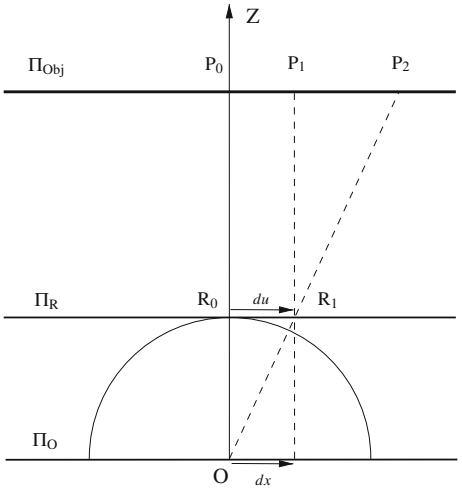


Fig. 5 Depth estimation using the positional and directional derivatives

Similarly we can calculate distance using the L_y, L_v values, thus in general:

$$Z_0 = \frac{L_u}{L_x} = \frac{L_v}{L_y}. \tag{10}$$

The distance value calculated from L_u, L_x should be the same as the distance value calculated from L_v, L_y. This is enforced by constraint (7).

3.5 Direct distance estimation using the equal triangles principle

Let us ignore the interpolation. Assuming the photoreceptors are arranged symmetrically, we could compute the distance from any of three orientations in the plane. For example, we could use ommatidia O^7, O^3 and O^6 (Fig. 3). This is shown next.

Let us take a look at a one-dimensional cross-section through the ommatidia in the model. Referring to Fig. 6, I denotes the intersection of two rays from neighboring ommatidia converging at an angle $\Delta\phi_{\mathrm{s}} = \Delta\phi_{\mathrm{r}} - \Delta\phi_{\mathrm{o}}$. I is at distance D_0 from the center of the eye and at distance D_1 from the object. A, B, C, D are points on the object projecting on photoreceptors R_7^7, R_7^6, R_6^6 and R_3^3. Let E_1 be the intensity difference between rays through A and D (we actually obtain the intensity difference between the rays through C and D, which is $2E_1$), and let E_2 be the intensity difference between the rays through A and B.

Then

$$\bar{CD} = 2\bar{AD} = 2D_1 \tan \Delta\phi_{\mathrm{s}} \quad \text{and}$$

$$\bar{AB} = (D_0 + D_1) \tan \Delta\phi_{\mathrm{o}}.$$

Taking the ratio and ignoring the tan, because angles $\Delta\phi_{\mathrm{s}}$ and $\Delta\phi_{\mathrm{o}}$ are sufficiently small, we have

$$\frac{\bar{AD}}{\bar{AB}} = \frac{E_1}{E_2} = \frac{D_1 \Delta\phi_{\mathrm{s}}}{(D_0 + D_1)\Delta\phi_{\mathrm{o}}}.$$

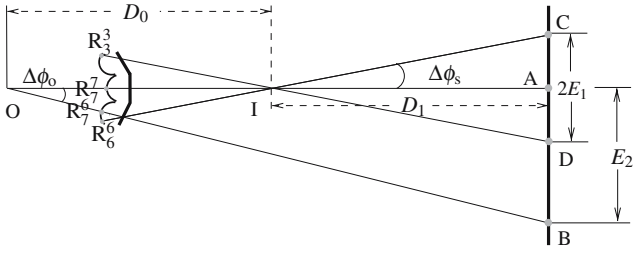


Fig. 6 Cross-section through three neighboring ommatidia

Through some manipulation we obtain

$$D_1 = \frac{D_0}{(\Delta\phi_s/\Delta\phi_o)(E_2/E_1) - 1}.$$

Thus the total distance D is

$$D = D_0 + D_1 = D_0 + \frac{D_0}{(\Delta\phi_s/\Delta\phi_o)(E_2/E_1) - 1}. \quad (11)$$

Notice that D_0 is a constant. Let us estimate its value. In our model the distance $R_7^7 R_6^6$ is approximately (5/4) $\Delta\phi_o$ (the distance from R_7^7 to R_7^6 is $\sim \Delta\phi_o$, the distance from R_7^6 to R_6^6 is $\sim (1/4)\Delta\phi_o$ and the three points approximately lie in a straight line). Thus, from the relation

$$\frac{5}{4}\Delta\phi_o = \Delta\phi_s(D_0 - 1)$$

by substituting

$$\Delta\phi_o = 1.88^\circ \text{ and } \Delta\phi_s = 0.3^\circ$$

we obtain

$$D_0 = \frac{5}{4}\frac{\Delta\phi_o}{\Delta\phi_s} + 1 \simeq 8.8. \quad (12)$$

So far, we have purposely avoided to mention the distance unit. In our model, we consider the sphere as having a radius of one unit, so every other distance is expressed in terms of the radius of our spherical camera. Having said that, if we consider a camera the size of an insect's eye i.e., radius equal to 0.5 mm, and an object that is 30 units away, then the actual distance of the object from the camera is 30×0.5 mm $= 15$ mm $= 1.5$ cm. Moreover the distance between the central points seen by the central photoreceptors of two adjacent ommatidia is $\Delta d \approx D \tan \Delta\phi_o \approx 0.5$ mm (when the plane is 15mm away). For such small distances, it is reasonable to assume that the intensity function obtained by the photoreceptors varies smoothly.

The results of the distance calculation for a number of different textures are presented in Sect. 5.

4 Neural network implementation

4.1 Artificial neural network

We first describe a simple artificial neural network approximating the computations described in Sects. 3.3 and 3.4. It consists of two networks in series; the first estimates the directional and positional derivatives by a simple linear system, the second one implements the depth computation as the ratio of the derivatives by a lookup table.

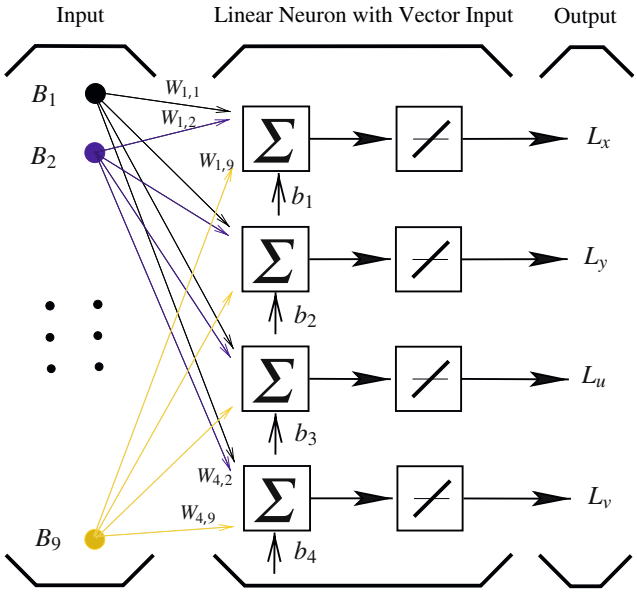


Fig. 7 Neural network computing the positional and directional derivatives. $\mathbf{B}_{9\times1}$ codes the intensity difference values obtained by the photoreceptors, $\mathbf{W}_{4\times9}$ is the weight vector, $\mathbf{b}_{4\times1}$ is the bias vector. The transfer function is the identity

In the first stage, the linear system (Eq.5) is solved with a single layered neural network using a linear transfer function, as shown in Fig. 7. The input to the network, denoted as $\mathbf{B}$ is a 9×1 vector which codes the differences in the intensity values recorded by the photoreceptors (the righthand side in Eq. 5). $\mathbf{W}$ is a 4×9 array that contains the weights of the neural network. Element W_{ij} specifies how the jth entry of $\mathbf{B}$ affects the ith unknown X_i. $\mathbf{b}$ contains the bias weights for each unknown. Vector $\mathbf{X}$ codes the x and y components of the directional and positional derivatives. The transfer function f is the identity function (i.e., $f(\mathbf{X}) = \mathbf{X}$). Training the neural network amounts to estimating the elements of $\mathbf{W}$ and $\mathbf{b}$. After that, vector $\mathbf{X}$ is computed as:

$$\mathbf{X} = f(\mathbf{WB} + \mathbf{b}) = \mathbf{WB} + \mathbf{b}. \quad (13)$$

In the second stage, depth is obtained as ratio of the directional and positional derivatives (Eq. 10). Instead of directly calculating the ratio, an approximate depth value is found using a *lookup table*. There are many possible neural network architectures approximating function mappings (Kohonen 2000). We use a *counterpropagation neural network* (CPN) (Hecht-Nielsen 1987, 1988), which can self-organize lookup tables. The basic idea is that, during adaptation, pairs of example vectors ($\mathbf{X}_e$, $\mathbf{Y}_e$) are presented to the network in the first and last layer. These vectors then propagate through the network in a counterflow manner to yield output vectors $\mathbf{X}$ and $\mathbf{Y}$, which are approximations of $\mathbf{X}_e$ and $\mathbf{Y}_e$. For our

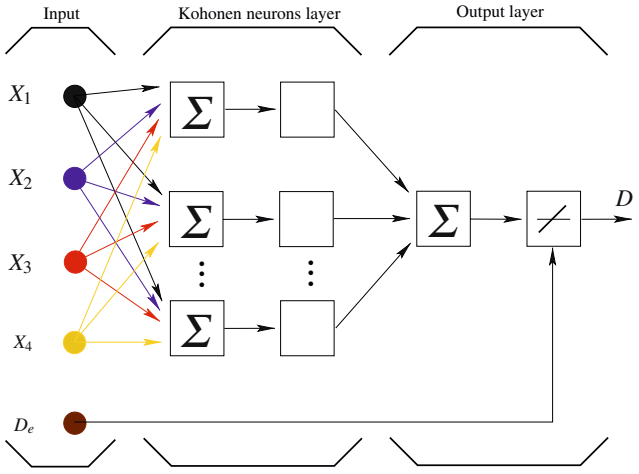


Fig. 8 The forward-only CPN implementing a lookup table for the estimation of distance as the ratio of directional and positional derivatives

purposes we use a simple variant of the CPN, which is known as forward-only counterpropagation neural network (see Fig. 8). It consists of one hidden layer, two input layers (with five neurons coding $X_1 - X_4$ and D_e) and one output layer (one neuron D). The input units $X_1 - X_4$ represent the four values of the positional and directional derivatives $(\frac{\partial L}{\partial x}, \frac{\partial L}{\partial y}, \frac{\partial L}{\partial u}, \frac{\partial L}{\partial v})$ and are connected with all the elements of the hidden layer. The hidden units compete with one another and a single "winner" emerges. Finally, the output unit D receives the output of the hidden units and produces as result the estimated depth. Neuron D_e is only connected with the output unit D, and it is used in the training stage only. It codes the true distance value, which could be provided from other sensor measurements. The size of the hidden layer greatly affects the accuracy of the estimated distance. From our experiments (Sect. 5) a layer of a few dozens neurons is expected to give reasonable results.

4.2 Discussion of biological implementation

It is well known that in the first stage the group of peripheral photoreceptors (R_1–R_6) terminate on the first optic neuropil (lamina) (Kirschfeld 1967), while the two central photoreceptors' axons bypass the lamina and project directly to the medulla (Campos-Ortega and Strausfeld 1972). In a second stage the signals from the peripheral photoreceptors (R_1–R_6) are conveyed from the lamina to the medulla by the L3 monopolar cells (Strausfeld 1989), where they associate with input from the central photoreceptors (Strausfeld 1984). There is also evidence of two independent, parallel pathways leading from medulla to lobula and lobula plate, respectively (Strausfeld and Lee 1991). It has been suggested that processing in the lobula plate is related to motion detection, while in the lobula the computation of form and color is performed (Bausenwein et al. 1992). Based on those findings the computations of depth, hypothesized here, would probably be realized in the lobula or in the medulla.

First, we want to point out that estimating the depth from the theory described here is just as complex as estimating image motion along the dominant gradient direction (the so-called normal flow Horn 1986). This becomes very clear when we compare the method to motion estimation techniques using image gradients. Such techniques require the spatial and temporal derivatives of the image intensity function to be computed; image motion is derived as the ratio of the temporal derivative over the spatial derivative. Other image motion models in the literature are based either on energy responses in the frequency domain or on correlation responses. They require finding the maximum energy response among a number of filter responses, or the maximum correlation response among a number of candidates. In addition, any method uses some form of interpolation. Depth estimation may be realized in a similar way. The derivatives as well as the ratio of derivatives may be derived using place coding (that is by choosing among a number of candidates) as opposed to value coding (that is by computing parametric values).

The network proposed above is simple enough to be implemented with only a few neurons. It is also quite robust, because by combining multiple intensity measurements in the first stage some noise and non-linearities can be tolerated. Alternatively, the depth may be estimated independently from the three directions of the photoreceptors, as described in Eq. (11). When computing depth from any individual direction, only the ratio of the intensity differences is required. Thus, this estimation could be approximated with a lookup table using a CPN in a way similar to the second network above. Following, the three depth estimates need to be combined. This can be realized by averaging, which amounts to lowpass filtering. The estimation of the image derivatives themselves can simply be implemented by a linear filter.

We have not mentioned the time dimension yet. The image intensities are also integrated over time, and this has the effect of additional low pass filtering. On the other hand, it could also be possible to defer the depth computations to the time domain. Image motion in the fly is modeled by the Reichardt detector (Reichardt 1969), which correlates two spatially and temporally

displaced signals. Estimating the motion amounts to choosing the time difference which best corresponds to the spatial difference. Using the motion hardware, instead of encoding depth through intensity differences, depth could be encoded as the time difference of Reichardt correlators. One set of Reichardt detectors could correlate the center rays, and another one the converging rays of neighboring ommatidia. The temporal time delay is inverse proportional to the spatial intensity change. Thus, the depth could be obtained, by modifying Eq. (11) to

$$D = D_0 + \frac{D_0}{(\Delta\phi_s/\Delta\phi_o)(T_1/T_2) - 1}, \tag{14}$$

where T_1 and T_2 denote the time difference between the peripheral and the central photoreceptors, respectively.

An important characteristic of the compound eye that we haven't discussed yet, is the angular sensitivity of the rhabdomeres (van Hateren 1984). Due to its small cross-section, incident light is channeled into the rhabdomere from a narrow range of directions, giving the photoreceptor a limited field of view. The spatial distribution function is called angular sensitivity and in the majority of compound eyes (including those of dipteran flies) is a triangular or bell-shaped function that is often approximated by a Gaussian curve (Götz 1964). Its shape is determined by the geometry and optical components of the ommatidium (Stavenga 2003a, 2004a). The half-width of the Gaussian is called the acceptance (or opening) angle $\Delta\rho$. It is found that the angular sensitivity depends on the wavelength of the incident light as well as its intensity. For low-light conditions $\Delta\rho$ is wider, while in bright light conditions it is narrower (Stavenga 2004b). More importantly, the angular sensitivity of the peripheral rhabdomeres (R_1–R_6) is distinctly larger than that of the central photoreceptors (R_7–R_8) (Stavenga 2003b). $\Delta\rho$ also depends on the species, with the ratio of acceptance angle to interommatidial angle commonly being one or less (Kirschfeld 1976).

In our model the intensity differences from the peripheral receptors are much smaller than those of the central receptors. Thus, it is critical that their estimates are quite accurate. A larger opening angle for those receptors and thus larger overlap of the visual fields in effect creates more low pass filtering. This should lead to a larger signal to noise ratio and thus better estimation. The signal to noise ratio generally is worse for low intensity values. A larger opening angle for decreasing light levels should make the depth estimation more stable at low light levels. Thus, in summary, the change in opening angle should have a positive effect on the depth estimation. But we do not expect that there is a significant difference in the accuracy of the depth estimation at different light levels because of the change in the size of the opening angle.

5 Experiments

5.1 Methodology

In the absence of a camera similar to the compound eye, we used the following methodology to test our algorithm. We gathered a number of images picturing real objects (e.g., trees, flowers, soil etc) as well as created a few images displaying special cases (e.g., linear, quadratic and cubic textures, horizontal and vertical edges, corners etc). We assume that those pictures are frontoparallel to the compound eye at various distances. Then we calculate the intensity value seen by each photoreceptor by intersecting its visual axis with the plane. We consider the angles $\Delta\phi_o$ and $\Delta\phi_r$ to be 1.88° and 2.18°, respectively, and use bilinear interpolation to get subpixel accuracy. We also assume that each photoreceptor has a field of view, described by an angular sensitivity function with acceptance angle $\Delta\rho = 1.9°$. Thus, we use a 2D Gaussian function whose size is determined by intersecting the plane with an appropriately oriented cone centered at the photoreceptor. As a consequence, the size of the Gaussian filter depends on the distance of the plane from the compound eye. The standard deviation of the function is also related to the size, more specifically we have chosen the standard deviation to be the same as the filter size. In the previous sections we presented two equivalent ways to estimate the distance, namely the non-linear optimization problem (Eq. 7) and the direct distance calculation using Eq. (11). Under ideal conditions (linear texture, no noise) both methods should give the same results. In practice the estimates are slightly, but not significantly different. Also note that the "equal-triangles" methods produces three estimates, one for each principal direction of the photoreceptors. Reasonable results are possible only if the data behaves linearly in the neighborhood considered. Thus, we chose the following simple implementation of a "linearity test". If the three distance values are close then we just average, otherwise we perform a linearity test on the intensity values. (That is, we check whether $L_7^j - L_7^7$ is close in value to $L_7^7 - L_7^{j+3}$ and $L_j^j - L_7^7$ is close in value to $L_7^7 - L_{j+3}^{j+3}$, $j \in \{1 \ldots 3\}$). Then we keep only the distance value corresponding to the directions which are better approximated as linear.

For each image and for distances ranging from 2 to 300 units away we apply both methods multiple times. Each time we assume that the central photoreceptor of

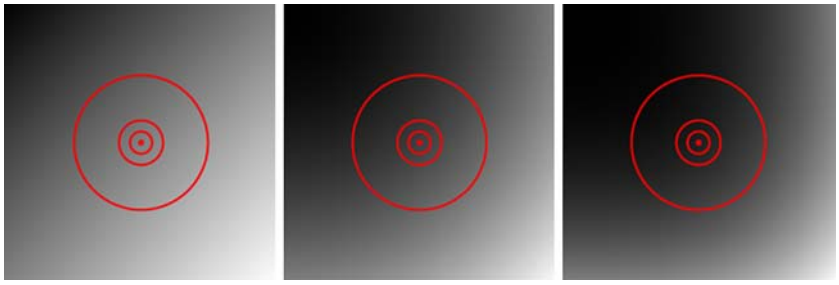


Fig. 9 The synthetic textures used in our experiments. From left to right the linear, quadratic and cubic texture is displayed. The four *circles* reveal the part seen by the seven ommatidia for distances 10, 50, 100 and 300 units away

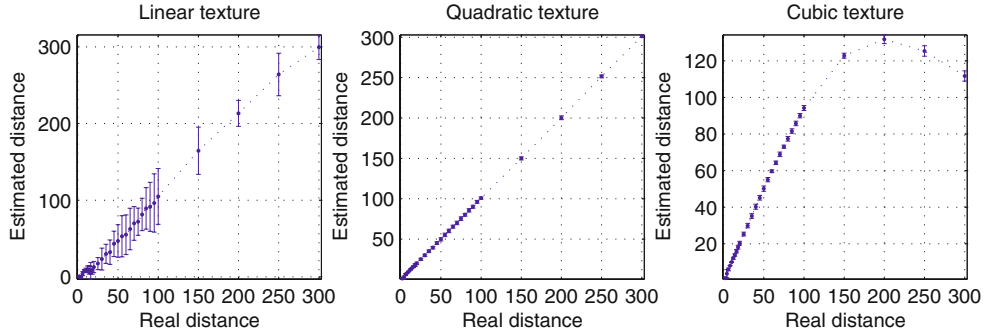


Fig. 10 The results of our experiments for linear, quadratic and cubic texture with noise using the non-linear optimization formulation. We use multiplicative noise with zero mean value and 2% standard deviation. We have taken 61 measurements for each texture (by considering a different center each time) and we present the median estimated distance and its interquartile range (IQR) divided by 2 using a dot and a vertical error bar, respectively. The effect of the multiplicative error is large in the case of the linear texture and very small for the other two textures

the central ommatidium sees a different pixel, thus we consider a different part of the picture. Then, we calculate a measure for the average value and its dispersion. In order to be robust in the presence of outliers we use the median value and the interquartile range i.e., the difference between the 75th and 25th percentile of the data. For simplicity, we only present the values obtained with the non-linear optimization problem. The results obtained with the other method are in most cases similar.

5.2 Results

Initially we use a plane with a linear, quadratic and cubic texture (Fig. 9). We obtain precise depth estimates for linear and quadratic textures. For cubic textures the results are within 1% of the real values for all but very large depth values (>1,000 units). Next, we superimpose multiplicative Gaussian noise with zero mean value and 2% standard deviation on the previous textures. The results are presented in Fig. 10. In this experiment distance estimation becomes more accurate as the real distance increases, which seems to contradict with the common sense. This phenomenon can be explained by the fact that the area seen by each photoreceptor increases as the texture is placed further away from the eye and as a consequence the effect of noise, which was added to the original image, is decreased. We elaborate more on the effect of noise on distance estimation later on this paper.

Next, we test the algorithm with real images.[5] Figure 11 presents a small but representative fraction of the images used. In order to simulate the limited acuity of each photoreceptor we filter each image with a circular averaging filter before applying our algorithm. The results are presented in Table 1.

According to that table as well as the outcome of our method with the rest of the images, depth can be estimated with good accuracy (<10% error) for close by (<20 units away) and with fair accuracy (∼25% error) for medium distances (<40 units away). More specifically from the table we observe that we got good distance estimates for the glass, medium for the rose and pavement and bad for the mud texture. The reason for the good results with the glass pattern is that this image has a good "linear" structure; going from dark on the lower left to bright on the upper right corner. So even though it contains significant energy at high spatial frequencies, which can be observed as fluctuations in the intensity

[5] Part of the images we used were from the collection cited in van Hateren and van der Schaaf [1998].

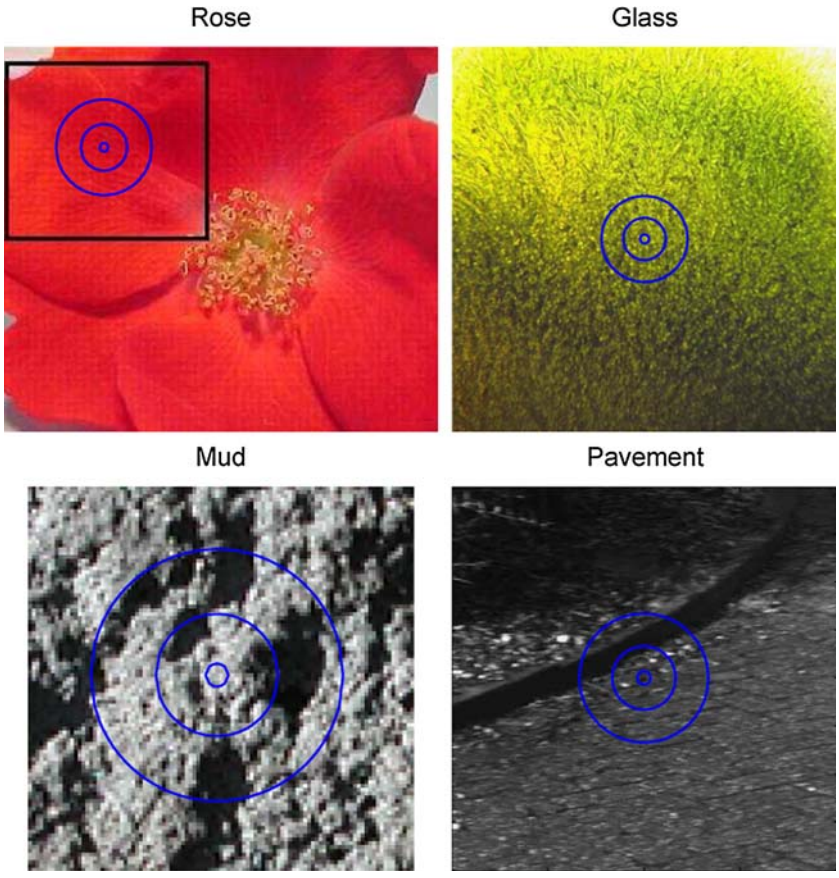


Fig. 11 Images of real objects used in our experiments. Starting from the *upper left image* and going clockwise: a rose, the microscopic structure of a glass, a pavement and a mud texture are depicted. Note that only the part of the rose inside the *black rectangle* was used in our experiments. Also note that the three *concentric circles* shown on each image reveal the area taken into account during the calculation of depth for distances of 10, 50 and 100 units respectively

Table 1 The results of our experiments for real images

True depth	Estimated depth			
	Rose texture	Glass texture	Mud texture	Pavement texture
4	3.87 ± 0.19	3.96 ± 0.12	3.94 ± 0.07	3.94 ± 0.11
6	5.98 ± 0.22	5.97 ± 0.11	5.93 ± 0.09	5.93 ± 0.17
8	8.00 ± 0.13	7.99 ± 0.05	7.94 ± 0.05	7.97 ± 0.07
10	10.08 ± 0.25	10.03 ± 0.10	10.17 ± 0.10	10.06 ± 0.14
12	12.30 ± 0.58	12.10 ± 0.24	12.82 ± 0.43	12.19 ± 0.35
14	14.53 ± 1.35	14.33 ± 0.50	16.04 ± 1.33	14.47 ± 1.04
16	16.64 ± 2.41	16.61 ± 0.81	20.68 ± 3.60	16.81 ± 1.67
18	18.48 ± 3.14	19.10 ± 1.39	25.38 ± 8.75	19.02 ± 2.57
20	20.76 ± 4.84	21.64 ± 1.54	28.09 ± 17.43	21.94 ± 3.09
25	26.97 ± 7.30	28.62 ± 2.10	17.75 ± 36.54	28.48 ± 6.41
30	34.58 ± 11.10	38.45 ± 6.74	-13.66 ± 31.76	33.31 ± 16.57
40	37.32 ± 27.12	56.31 ± 15.16	-8.52 ± 19.49	26.26 ± 42.73
50	46.01 ± 28.30	70.52 ± 11.78	-8.80 ± 4.91	20.13 ± 46.45
60	57.05 ± 80.45	75.37 ± 17.67	-6.13 ± 1.54	17.46 ± 40.08
70	-59.67 ± 71.23	80.18 ± 51.87	-4.44 ± 1.50	8.42 ± 37.69
80	-45.48 ± 22.92	94.87 ± 86.92	-3.34 ± 1.15	-8.08 ± 35.04
90	-32.41 ± 15.46	108.83 ± 80.24	-2.17 ± 1.35	-9.41 ± 28.44
100	-23.21 ± 10.77	131.69 ± 64.44	-1.02 ± 1.26	-10.40 ± 30.69

The sizes for the rose, glass, mud and road texture used are $120 \times 120, 600 \times 600, 444 \times 500$ and $1,024 \times 1,024$ pixels, respectively. At a distance of 10 units away from the camera the distance between two outer opposite ommatidia centers i.e., R_7^6 and R_7^3 (referring to Fig. 3) is 6.6 pixels for the rose and mud image and 13.2 pixels for the other images. Similarly, for depths 50 and 100 units the distance is 32.8, 65.6 and 65.6, 131 pixels, respectively. We have taken 61 measurements for each image (by considering a different center each time) and we display the median estimated distance and its interquartile range divided by 2

values of neighboring pixels, after the low-band filtering step the "linear" structure emerges and makes distance estimation accurate. The two patterns, glass and pavement, have dominant low frequencies. As we explain in the next section, this fact makes depth estimation at larger distances less accurate. Finally, the mud pattern contains significant medium and high frequencies and doesn't have the linear structure of the glass image. This is the most unreliable case for our algorithm.

5.3 Analysis of the distance estimation

The method is based on the linearity assumption, namely that the scene observed by the part of the compound eye has linear texture. In practice this assumption may not hold, resulting in poor distance estimation. In this section we explain why the algorithm is not accurate for medium and large distances. We also take a look at the signal in the frequency domain to investigate the effect of dominant low and high frequencies on distance estimation.

As it is derived in Eq. (11), the analytical relation between distance and image intensity is

$$D = D_0 + D_1 = D_0 + \frac{D_0}{(\Delta\phi_s/\Delta\phi_o)(E_2/E_1) - 1}.$$

If we consider the ratio E_2/E_1 as a single variable x then the above equation is a hyperbolic function

$$D(x) = \alpha + \frac{\alpha}{\beta x - 1}, \quad \text{where} \quad \alpha = D_0, \quad \beta = \frac{\Delta\phi_s}{\Delta\phi_o}, \quad x = \frac{E_2}{E_1}.$$

Figure 12 presents the plot of $D(x)$. It is clear from the graph that for distances greater than 25 units good measurements of E_2/E_1 are needed, as a small error in this ratio will result in a large error in distance estimation.

To backup the previous argument, in Fig. 13 we present the part of the pavement texture seen by the set of seven ommatidia at various distances between 18 and 50 units. We consider the three principal directions by connecting antidiammetric diamonds along the periphery of the large circle. Bottom-left to top-right is direction one, top-left to bottom-right is the second, and top to bottom is the third. According to our experiments, depth estimation is quite accurate (less than 20% error) for all distances along the first direction, accurate for distances up to 25 units away for the third and up to 20 units away for the second one. Those results can be explained as follows; the first direction is almost the same as the direction of the gradient, thus the change on the intensity values is significant, while on the remaining directions the intensity change is small and even small deviations from linearity result in bad distance estimation. Especially the second direction has almost uniform intensity, that's why the results along this direction are worse. Also note that the results are getting worse as the distance of the texture from the eye is increased because distance estimation is a hyperbolic function.

Let us now take a look at how the frequency spectrum of the image affects the error in the distance estimation.

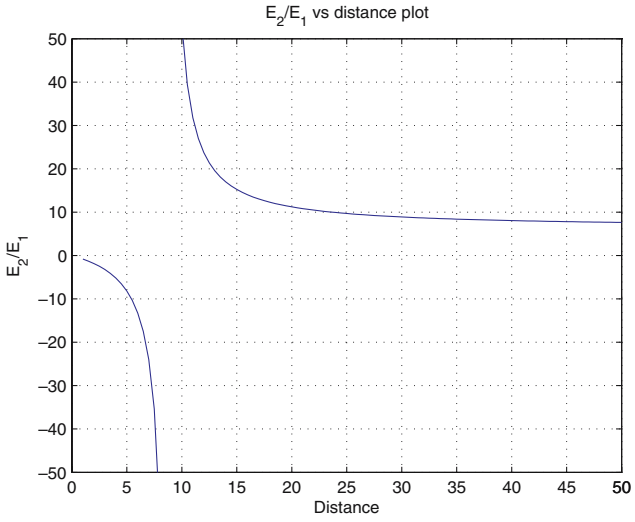


Fig. 12 Distance as a function of E_2/E_1

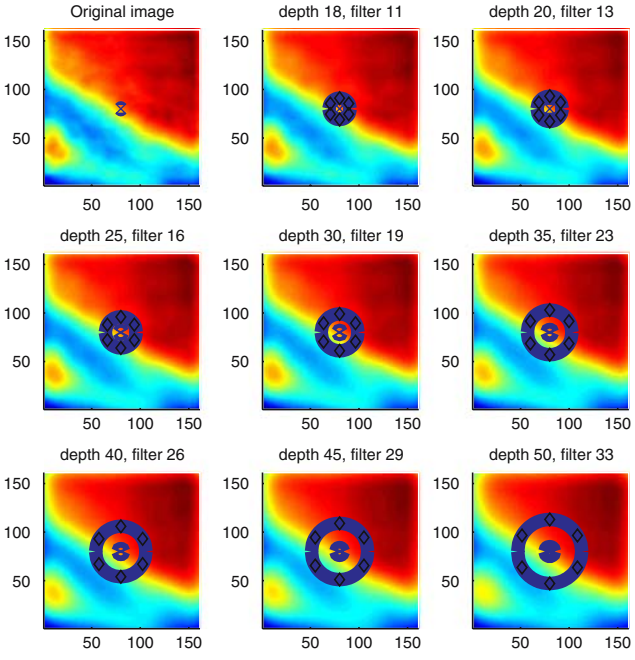


Fig. 13 Part of the pavement texture seen by a set of seven neighboring ommatidia. The *large* and *small circles* point at the central image points observed by the central and "converging" photoreceptors, respectively. The image is displayed after a Gaussian filter of appropriate size has been applied. Hence, on the electronic version of the document one can make out that the image is getting smoother as the distance increases, because the Gaussian filter is also getting larger

If the image is "flat", meaning that the intensity differences (E_1, E_2) are small, a small error in the intensity estimates can lead to a large error in the estimate of the ratio E_2/E_1, resulting in a large error in distance estimation. Thus, the intensity function should not contain only low frequencies. The "rose" texture is an example of a texture containing dominant low frequencies, and the distance estimation results are not very good for that pattern.

As the images should not be too flat, it is also true that the images should not contain dominant high frequencies. Let the image signal $s(x)$ be a one-dimensional sinusoid of frequency ω i.e., $s(x) = \sin(\omega x)$, and let this signal be observed by a central and two peripheral antidiammetric ommatidia, which are at distance W (in our model $W = 2\Delta\phi_o \simeq 0.066$ units, with $\Delta\phi_o$ being the interommatidial angle). Ideally, the signal should be monotonically increasing. On average, this is the case when the distance between the peripheral photoreceptors is 1/8 of the period of the signal. This imposes the following constraint on the highest frequency ω_{max} of the signal $s(x)$.

$$\omega_{max} W \leq \frac{\pi}{4} \Rightarrow \omega_{max} \leq \frac{\pi}{4W} = \frac{\pi}{8\Delta\phi_o}. \quad (15)$$

The signal observed by a photoreceptor is filtered by a Gaussian kernel whose cutoff frequency ω_{cutoff} is about

$$\omega_{cutoff} = \frac{1}{2(\Delta\rho/2)} = \frac{1}{\Delta\rho}. \quad (16)$$

With $\Delta\phi_o$ and $\Delta\rho$ close in value, the cutoff frequency is about twice as large as the maximum tolerable frequency. Thus, in general some high frequencies in the intensity signal of the light field remain after the filter is applied. Furthermore, since the truncated Gaussian is not an ideal low-pass filter, there is a problem in the case of very strong high frequencies, as we observed for the random noise in the first experiment as well as the mud texture. For natural images, however, on average, lower frequencies are more dominant, with the power spectrum falling approximately inverse proportional to the square of the spatial frequency (Simoncelli and Olshausen 2001). Thus, in most cases high frequencies should not present a problem for the flies.

6 Conclusions

In this paper we presented a theory on how with the compound eyes of dipteran flies distance can be estimated. Experiments based on images of objects along with an analysis indicate that distance estimation is possible with good accuracy if the viewed objects are relatively close to the eye. If such a mechanism is used by the fly, it thus would be useful in walking conditions.

The idea underlying the theory is to exploit the differential structure of light. Another way of looking at the approach is as a form of differential stereo. In regular stereo with two parallel cameras, one has to match points in the two images. The distance of the matching points defines the image disparity. In our approach the image disparity d is approximated linearly from the intensity derivatives L_x between parallel rays and the intensity derivatives L_u of converging rays; i.e., $d = L_x/L_u$. The estimated distance, then, is inversely proportional to the disparity; i.e., $Z_0 = 1/d = L_u/L_x$.

We should point out that the analysis here relies on the data of Pick [1977]. However, the method is also possible (for small distance values, if, as it is often modeled in other approaches, the neurally superimposed rays were parallel and not converging (as in Pick's data). In such case the estimation of the derivatives becomes a bit simpler as shown in Sect. 3.3 Eq. (8).

Another issue is whether this mechanism could aid in motion estimation, which is most essential for insects. If 3D structure estimates are available, and even if they are ordinal only, their knowledge could certainly facilitate the process of segmentation (Brodský et al. 1999). Distance estimates theoretically could also be used to facilitate the estimation of ego-motion. The essential idea is that the introduction of depth information makes the estimation of 3D motion linear (Neumann et al. 2004). However, for flying conditions the depth estimates by our model are expected to be very erroneous. Thus, the resulting ego-motion estimation should not be accurate enough to be useful. Studies with bees also indicate, that bees with similar eyes do not use depth estimates in conjunction with motion in flying conditions. For example, Srinivasan et al. [1989] has demonstrated that during locomotion bees use directly the size of the flow. A moving wall caused bees to change their distance to the wall (Srinivasan et al. 2000). The distance a bee flies seems to be encoded in the amount of flow on the eye. Thus, bees were fooled in their judgements when flying through a tunnel.

References

Adelson E, Bergen J (1991) The plenoptic function and the elements of early vision. In: Computational models of visual processing. MIT, Cambridge, pp 3–20

Anderson JC, Laughlin SB (2000) Photoreceptor performance and the co-ordination of achromatic and chromatic inputs in the fly visual system. Vis Res 40:13–31

Barlow H (1964) The physical limits of visual discrimination. In: Giese A (ed) Photophysiology. Academic, New York, pp 163–202

Bausenwein B, Dittrich AP, Fischbach KF (1992) The optic lobe of Drosophila melanogaster. II Sorting of retinotopic pathways in the medulla. Cell Tissue Res 267:17–28

Bolles RC, Baker HH, Marimont DH (1987) Epipolar-plane image analysis: an approach to determining structure from motion. Int J Comp Vis 1:7–55

Braitenberg V (1967) Patterns of projection in the visual system of the fly. Exp Brain Res 3:271–298

Brodský T, Fermüller C, Aloimonos Y (1999) Shape from video. Proc IEEE Conf Comput Vis Pattern Recognit 2:146–151

Campos-Ortega J, Strausfeld N (1972) The columnar organisation of the second synaptic region of the visual system of Musca domestica. Z Zellforschung 124:561–585

Dahmen H, Wüst RM, Zeil J (1997) Extracting egomotion parameters from optic flow: Principal limits for animals and machines. In: Srinivasan MV, Vekatesh S (eds) From living eyes to seeing machines, Oxford University, Oxford, pp 174–198

Dawkins R (1996) Climbing mount improbable. Norton, New York

Fermüller C, Aloimonos Y (2000) Observability of 3D motion. Int J Comput Vis 37:43–63

Franceschini N (1975) Sampling of the visual environment by the compound eye of the fly. In: Snyder AW, Menzel R (eds) Photoreceptor optics. Springer, Berlin Heidelberg New York, pp 98–125

Gortler S, Grzeszczuk R, Cohen RSM (1996) The lumigraph. ACM SIGGRAPH, pp 43–54

Götz K (1964) Optomotorische Untersuchung des visuellen Systems einiger Augenmutanten der Fruchtfliege Drosophila. Kybernetik 2:77–92

van Hateren JH (1984) Waveguide theory applied to optically measured angular sensitivities of fly photoreceptors. J Comp Phys A 154:761–771

van Hateren JH, van der Schaaf A (1998) Independent component filters of natural images compared with simple cells in primary visual cortex. Proc R Soc Lond B 265:359–366

Hecht-Nielsen R (1987) Counterpropagation networks. Appl Optics 26:4979–4984

Hecht-Nielsen R (1988) Applications of counterpropagation networks. Neural Netw 1:131–139

Horn BKP (1986) Robot vision. McGraw Hill, New York

Kirschfeld K (1967) Die Projektion der optischen Umwelt auf das Raster der Rhabdomere im Komplexauge von Musca. Exp Brain Res 3:248–270

Kirschfeld K (1976) The resolution of lens and compound eyes. In: Zettler Z, Weiler R (eds) Neural principles in vision. Springer, Berlin Heidelberg New York, pp 354–370

Kohonen T (2000) Self-organizing maps. Springer, Berlin Heidelberg New York

Land M (1981) Optics and vision in invertebrates. In: Autrum H (ed) Comparative physiology and evolution of vision in invertebrates. Handbook of sensory physiology, vol VII/6B. Springer, Berlin Heidelberg New York, pp 471–592

Laughlin SB (1981) Neural principles in the peripheral visual systems of invertebrates. In: Autrum H (ed) Comparative physiology and evolution of vision in invertebrates. Handbook of sensory physiology, vol VII/6B. Springer, Berlin Heidelberg New York, pp 133–280

Levoy M, Hanrahan P (1996) Light field rendering. ACM SIGGRAPH, pp 161–170

Neumann J, Fermüller C (2003) Plenoptic video geometry. Vis Comput 19:395–404

Neumann J, Fermüller C, Aloimonos Y (2004) A hierarchy of 3D photography. Comput Vis Image Underst 96:274–293

Pick B (1977) Specific misalignments of rhabdomere visual axes in the neural superposition eye of dipteral flies. Biol Cybern 26:215–224

Reichardt W (1969) Movement perception in insects. In: Reichardt W (ed) Processing of optical data by organisms and by machines. Academic, New York

Rossel S (1983) Binocular stereopsis in an insect. Nature 301: 821–822

Simoncelli EP, Olshausen BA (2001) Natural image statistics and neural representation. Annu Rev Neurosci 24:1193–1216

Srinivasan MV, Lehrer M, Zhang SW, Horridge GA (1989) How honeybees measure their distance from objects of unknown size. J Comp Physiol A 165:605–613

Srinivasan MV, Zhang SW, Altwein M, Tautz J (2000) Honeybee navigation: nature and calibration of the 'odometer'. Science 287:851–853

Stavenga DG (2003a) Angular and spectral sensitivity of fly photoreceptors. I. Integrated facet lens and rhabdomere optics. J Comp Physiol A 189:1–17

Stavenga DG (2003b) Angular and spectral sensitivity of fly photoreceptors. II. Dependence on facet lens F-number and rhabdomere type in Drosophila. J Comp Physiol A 189:189–202

Stavenga DG (2004a) Angular and spectral sensitivity of fly photoreceptors. III. Dependence on the pupil mechanism in the blowfly Calliphora. J Comp Physiol A 190:115–129

Stavenga DG (2004b) Visual acuity of fly photoreceptors in natural conditions-dependence on UV sensitizing pigment and light-controlling pupil. J Exp Biol 207:1703–1713

Strausfeld NJ (1984) Functional neuroanatomy of the blowfly's visual system. In: Ali MA (ed) Photoreception and vision in invertebrates. Plenum Press, New York, pp 483–522

Strausfeld NJ (1989) Beneath the compound eye: neuroanatomical analysis and physiological correlates in the study of insect vision. In: Stavenga DG, Hardie RC (eds) Facets of vision. Springer, Berlin Heidelberg Newyork, pp 317–359

Strausfeld NJ, Lee JK (1966) Neuronal basis for parallel visual processing in the fly. Vis Neurosci 7:13–33

Trujillo-Cenoz O, Melamed J (1966) Electron microscope observations on the peripheral and intermediate retinas of dipterans. In: Bernard CG (ed) The functional organization of the compound eye. Pergamon Press, Oxford, pp 339–361