Correlative velocity estimation: visual motion analysis, independent of object form, in arrays of velocity-tuned bilocal detectors

Helmut Glünder

Institut für Medizinische Psychologie, Ludwig-Maximilians-Universität München, Gœthestraße 31, D-8000 München 2, Federal Republic of Germany

Received May 3, 1989; accepted September 19, 1989

The visual estimation of object velocity in systems of tuned bilocal detector units (simplified Hassenstein–Reichardt detectors) is investigated. The units contain delay filters of an arbitrary low-pass characteristic. Arrays of such detector units with identical delay filters are assumed to cover the plane of analysis. The global evaluation of the output signals of suitably arranged detector units is exemplified by the analysis of frontoparallel translations of rigid objects. The correlative method permits the estimation of the instantaneous object velocity, independently of object form. The time course of the resulting estimate is shown to be the convolution of the true velocity profile with a time-invariant kernel that depends solely on the impulse response of the delay filters and thus characterizes the analyzer system. The mathematical analysis of the processing principle is illustrated by considering idealized detector systems. The response of correlative motion analyzers to compound motion and to motion of nonrigid objects is discussed.

1. INTRODUCTION

There is once again considerable interest in a detector principle for the visual analysis of motion,[1–5] which was originally proposed by Hassenstein and Reichardt in the 1950's.[6–9] This so-called bilocal correlation approach was originally formulated in order to explain the optomotor behavior of invertebrates and the corresponding neural structures in their compound eyes.[10–15] Meanwhile, however, there is evidence that this correlation principle also applies to motion analysis in vertebrates, including humans.[1,16–18]

The function of bilocal velocity detectors is based on the measurement of the time τ that is needed by a moving object point to traverse a spatial interval of known length ρ. The estimated velocity of the point is then given by the quotient

$$v = \rho/\tau. \tag{1}$$

The measurement of time can be circumvented by comparing the object velocity with a set of velocity prototypes that are implemented by elementary detectors, each tuned to a different velocity. Figure 1a shows a possible schematic of such a detector unit. More elaborated versions and neurobiologically inspired models of this correlation detector are discussed by van Santen and Sperling[2] as well as Ruff *et al.*[19] The detector unit considered here senses the image intensity at two positions separated by a distance ρ. In the simplest case, one of the intensity signals thus determined is delayed by an amount τ at stage TF and is compared with the other signal at stage NL. Coincidence of the two signals is obtained if an object point crosses the sensors in the correct succession and at a speed according to Eq. (1). However, when confronted with more-complex stimuli, a single detector unit of this type in general gives rise to object-specific deviations from this ideal result. In this paper it is proposed to overcome this object dependence of the velocity measurements by using arrays of such elementary detectors and by suitably combining their coincidence signals. Velocity tuning of the bilocal unit is obtained by the proper choice of the spatiotemporal intervals. For the human visual system van de Grind *et al.* recently found a constant time delay for a wide range of velocities (at least for $2°/\text{sec} < v_{\text{foveal}} < 30°/\text{sec}$).[17] This implies that tuning takes place by alterations of the sensor spacing ρ. The following investigations deal with arrays of such ρ-tuned elementary detectors situated in the image plane of an optical system (retina) or in a preprocessed but still retinotopically organized image representation.

In this paper I address issues that are exemplified by the following questions on motion analysis in two-dimensional arrays of bilocal detectors[2–4,20–22]:

- How should the coincidence or output signals of the detector units be combined?
- How does object form affect the measurement of velocity in an analyzer system?
- How does an analyzer system respond to velocity transients?

In what follows it is demonstrated that the instantaneous velocity of an object can be estimated independently of its form and without any additional measurements of form parameters, such as curvature. This is achieved by suitably integrating and evaluating the coincidence signals of bilocal units that are properly arranged in the detector plane. The distribution of the units is specific for each type of motion, e.g., frontoparallel translation or rotation, expansion, and rotation in depth. In a first step the output signals of elementary detectors with identical geometric parameters are pooled. This yields the so-called modified autocorrelation function. In a second step the maximum of this function, or

0740-3232/90/020255-09$02.00 © 1990 Optical Society of America

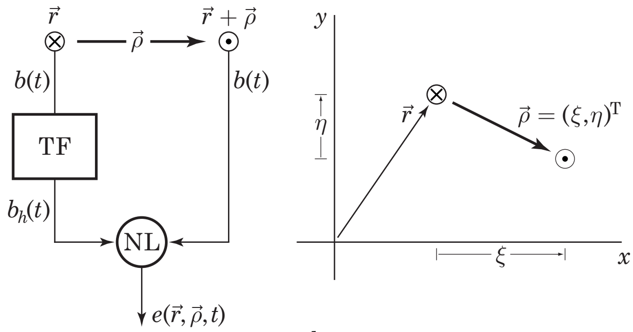

Fig. 1. a, Schematic of a velocity-tuned bilocal detector unit. TF indicates a linear, time-invariant filter of impulse response $h(t)$, and NL indicates a nonlinear operation (e.g., multiplication) of two time-varying intensity signals. b, Location of a unit's point sensors in the detector plane.

alternatively its centre of gravity (centroid), is determined. The position vector thus computed is proportional to the desired instantaneous velocity estimate. For the centroid-based approach, the estimated velocity as a function of time is proved to be the convolution product of the true velocity profile with a time-invariant kernel that is characteristic of the analyzer system. The kernel function depends on the integral function of the impulse response of the (delay) filter, which is part of each detector unit. It is shown that any linear temporal filter with an impulse response of non-vanishing mean may serve for this purpose.

The approach to form-independent velocity estimation is explicated for arbitrary frontoparallel translations, although it applies equally well to other types of rigid object motion in space. This choice was made in order to permit comparisons with other investigations that concern mostly this type of motion. For reasons of simplicity the mathematical treatment is confined to the global analysis of a single and rigid object translating in front of an unstructured background. In order to meet real-world requirements, local or regional analyses must be assumed to be operative in biological systems.[5,23,24] Unlike other approaches, the method proposed here is not restricted to the analysis of (temporally discrete) image sequences.

2. ANALYZER FOR FRONTOPARALLEL TRANSLATORY MOTION

A system is considered that estimates the velocity $\mathbf{v}(t) = [v_x(t), v_y(t)]^{\mathrm{T}}$ of a rigid object function $b(x,y) =: b(\mathbf{r}) \neq$ const. that is arbitrarily translating in an otherwise unstructured image plane. The front end of this analyzer system is assumed to be a spatially continuous arrangement of bilocal motion detector units in this plane. One of these units is sketched in Fig. 1a. It represents one velocity-tuned branch of the classical Hassenstein–Reichardt detector. This detector is characterized by the vector $\boldsymbol{\rho}$, which specifies the relative position of its sensor points in the image plane, i.e., their spatial separation and orientation (cf. Fig. 1b), as well as the preferred direction for a moving point. It is further typified by the impulse response $h(t)$ of a causal temporal filter situated in one branch of the detector circuit. The output signal $e(\mathbf{r},\boldsymbol{\rho},t)$ of such a bilocal detector results from the nonlinear combination of the filtered input signal $b_h(t)$ at location $\mathbf{r}$ with the direct input signal $b(t)$ at point $\mathbf{r}+\boldsymbol{\rho}$ of the image plane. For the following mathematical analyses, the multiplication is chosen as the nonlinear operation, and all motion-detector units of the analyzer system are assumed to have identical filters. Furthermore, the detector units are assumed to be isotropically arranged at every image point, i.e., their vectors $\boldsymbol{\rho}$ cover the whole range of angles ($0° \leq \vartheta < 360°$) and the whole range of radial spans ($0 < \rho \leq \rho_{\max}$).[16] The upper limit $\rho_{\max}$ depends on the size of the region of interest and on the maximum speed to be analyzed (see Section 7).

For the global analysis of a translating object function $b(\mathbf{r})$ it is suggested that the output signals $e(\mathbf{r},\boldsymbol{\rho},t)$ of all detector units with identical vectors $\boldsymbol{\rho}$ be separately integrated. This pooling scheme is mathematically described as correlation of the moving object function $b(\mathbf{r},t)$ with its temporally filtered version $b_h(\mathbf{r},t)$. Hence the resulting function is called a modified autocorrelation function (MACF) of order 2:

$$\begin{aligned} k(\boldsymbol{\rho},t) &= \iint e(\mathbf{r},\boldsymbol{\rho},t)\,\mathrm{d}\mathbf{r} = \iint b(\mathbf{r}+\boldsymbol{\rho},t)\,b_h(\mathbf{r},t)\,\mathrm{d}\mathbf{r} \\ &=: b(\mathbf{r},t) \circledast^{\rho} b_h(\mathbf{r},t) \\ &\qquad \text{for real-valued } b(\mathbf{r},t) \text{ and } b_h(\mathbf{r},t). \end{aligned} \tag{2a}$$

Unless otherwise stated, all integrations in this paper extend over the support of the integrands. The symbol $\circledast$ as specified by the (shift) variable $\boldsymbol{\rho} = (\xi,\eta)^{\mathrm{T}}$ indicates the two-dimensional spatial correlation.

3. PRINCIPLE OF VELOCITY ESTIMATION

The question arises: How can estimates of the true object velocity $\mathbf{v}(t)$ be gained from the MACF $k(\boldsymbol{\rho},t)$? In considering the circuit of Fig. 1a it becomes immediately clear that it is impossible to compute velocity estimates $\mathbf{v}_h(t)$ if detector units without filters are applied, i.e., for filters with the impulse response

$$h_z(t) \equiv \delta(t) \;\Rightarrow\; \mathbf{v}_z(t) \equiv \mathbf{0}. \tag{3}$$

One might predict that the true velocity can be computed if filters are used that introduce an infinitesimal signal delay:

$$h_\varepsilon(t) = \lim_{\tau_\varepsilon \to 0} \delta(t-\tau_\varepsilon) \;\Rightarrow\; \mathbf{v}_\varepsilon(t) = \mathbf{v}(t). \tag{4}$$

For an ideal signal delay of finite delay time τ_0 the true velocity can be approximated by the quotient of spatial and temporal intervals [cf. Eq. (1)]:

$$h_0(t) = \delta(t-\tau_0) \;\Rightarrow\; \mathbf{v}_0(t) = \boldsymbol{\rho}_0(t)/\tau_0\,. \tag{5}$$

Vector $\boldsymbol{\rho}_0(t)$ indicates the true translation of the object from the position at a time interval τ_0 in the past to its present location. This vector can always be found from the coordinates of the absolute maximum of the MACF $k_0(\boldsymbol{\rho},t)$:

$$k_0(\boldsymbol{\rho}_0,t) := \max_{\rho}[k_0(\boldsymbol{\rho},t)], \tag{6}$$

provided that the object function is nonperiodic. Thus the estimated velocity $\mathbf{v}_0(t)$ is independent of the object function. Motion analysis performed on (temporally discrete) image sequences, e.g., in computer vision, is a special case, because time t is discrete and especially because the temporal increment in general serves as delay time τ_0. Unlike technical systems, biological systems must deal with continuous time t and band-limited filters, implying extended impulse responses. In Section 5 it is demonstrated how velocity estimates can be obtained from MACF's that are computed by such systems.

4. DESCRIPTION OF TRANSLATORY IMAGE MOTION IN THE SPACE–TIME DOMAIN

A point situated at the origin of a translating pictorial object leaves a trace that can be mathematically described as a δ-line function in space-time. The three-dimensional extension of Papoulis' formulation[25] leads to a definition of δ-line functions as lines of intersection (product) between two δ surfaces. Mainly because of causality, the δ-line functions considered here are unique with respect to the temporal coordinate. Therefore only one spatial object position is possible at any moment. Consequently, the two δ surfaces are defined by argument functions $a(\mathbf{r}, t)$ of one spatial coordinate:

$$a_x(x,t) = x - X(t) \tag{7a}$$

and

$$a_y(y,t) = y - Y(t). \tag{7b}$$

The product of the corresponding δ surfaces describes the point trace in the xyt space:

$$\delta[\mathbf{r} - \mathbf{R}(t)] := \delta[a_x(x,t)]\,\delta[a_y(y,t)] = \delta[x - X(t)]\,\delta[y - Y(t)], \quad \text{with } \mathbf{R}(t) = [X(t), Y(t)]^{\mathrm{T}}. \tag{8}$$

Vector $\mathbf{R}(t)$ is the trajectory of the moving object in the xy plane as a function of time. Its temporal derivative is its true velocity $\mathbf{v}(t)$. Object motion, i.e., the spatiotemporal pattern $b(\mathbf{r}, t)$, can now be expressed as the convolution of the object function $b(\mathbf{r})$ with this δ-line function (cf. Appendix A):

$$b(\mathbf{r},t) = b(\mathbf{r})\,\delta(t) \ast \delta[\mathbf{r} - \mathbf{R}(t)] = b[\mathbf{r} - \mathbf{R}(t)]. \tag{9}$$

Unless otherwise specified, the symbol $\ast$ denotes the three-dimensional spatiotemporal convolution.

With the help of Eq. (9) the temporally filtered version $b_h(\mathbf{r}, t)$ of the spatiotemporal pattern $b(\mathbf{r}, t)$ can now be formulated more explicitly [cf. Eq. (2a)]:

$$b_h(\mathbf{r},t) = b(\mathbf{r},t) \ast h(t)\,\delta(\mathbf{r}) = b(\mathbf{r})\,\delta(t) \ast \delta[\mathbf{r} - \mathbf{R}(t)] \ast h(t)\,\delta(\mathbf{r}), \quad \text{with } \delta(\mathbf{r}) := \delta(x)\,\delta(y). \tag{10}$$

The δ-line function $\delta[\mathbf{r} - \mathbf{R}(t)]$ determines not only the spatiotemporal position of the filter impulse response $h(t)$ but, depending on the reciprocal of the true speed, also its magnitude [for details see Appendix A; the case $v(t) = 0$ is discussed in Section 7].

By using Eqs. (9) and (10), the MACF of Eq. (2a) can now be formulated in a more detailed way:

$$k(\boldsymbol{\rho},t) = \{b(\mathbf{r})\,\delta(t) \ast \delta[\mathbf{r} - \mathbf{R}(t)]\} \circledast^{\rho} \{b(\mathbf{r})\,\delta(t) \ast \delta[\mathbf{r} - \mathbf{R}(t)] \ast h(t)\,\delta(\mathbf{r})\} \tag{2b}$$

and rearranged as follows:

$$k(\boldsymbol{\rho},t) = \{[b(\mathbf{r}) \circledast^{\rho} b(\mathbf{r})]\,\delta(t)\} \ast \big(\delta[\mathbf{r} - \mathbf{R}(t)] \circledast^{\rho} \{\delta[\mathbf{r} - \mathbf{R}(t)] \ast h(t)\,\delta(\mathbf{r})\}\big). \tag{2c}$$

Introducing the static autocorrelation function (ACF) $k_a(\boldsymbol{\rho})$ of the object function $b(\mathbf{r})$ permits a compact formulation of the MACF:

$$k(\boldsymbol{\rho}, t) =: k_a(\boldsymbol{\rho}) \ast^{\rho} u(\boldsymbol{\rho}, t). \tag{2d}$$

The symbol $\ast$ as specified by the (shift) variable $\boldsymbol{\rho} = (\xi, \eta)^{\mathrm{T}}$ denotes the two-dimensional spatial convolution.

Equations (2c) and (2d) reveal an important property of the MACF: It can be expressed as the convolution of the motion-independent ACF $k_a(\boldsymbol{\rho})$ with an object-independent function $u(\boldsymbol{\rho}, t)$. Whereas the former typifies the object, the latter is the system-specific motion function (SMF), which depends solely on the motion history and on the impulse response of the filter. The SMF contains all the information about motion that can be supplied by an analyzer system of the type discussed. This information is at least theoretically available for the estimation of the instantaneous object velocity, but it is accessible only through the MACF. Finally, it follows from Eqs. (10) and (A5) below that the spatial mean of the SMF equals the temporal mean of the filter impulse response for every moment:

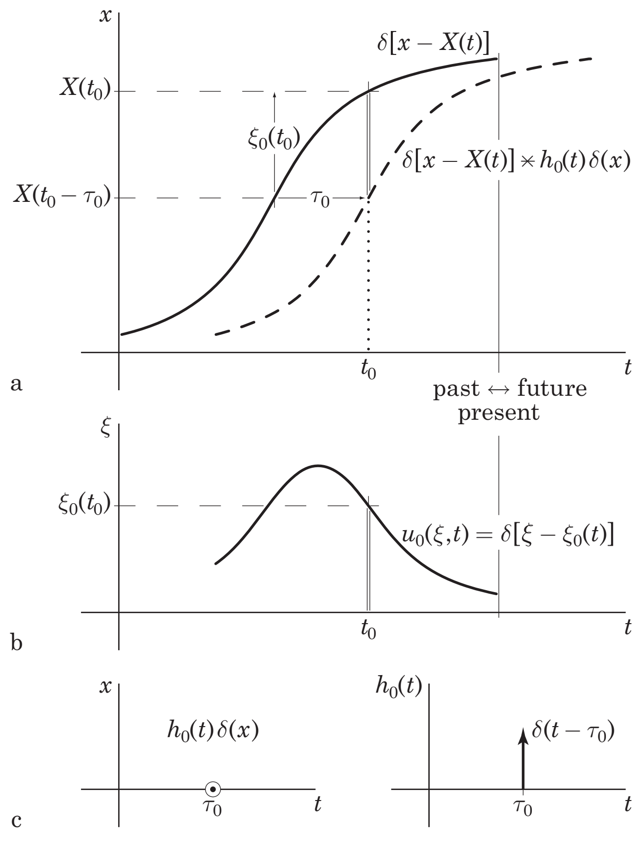

Fig. 2. a, One-dimensional motion along the x coordinate described by the trajectory $[x - X(t)]$ in the space-time domain. A correlative analyzer system composed of detector units with identical filters of delay time τ_0 is considered for velocity estimation. b, The SMF of the motion trajectory shown in a. The evaluation of the instantaneous velocity $v_0 = \xi_0/\tau_0$ at time t_0 is indicated. c, The impulse response of the ideal delay filter.

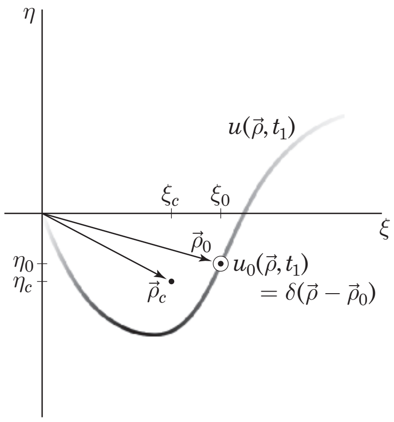

Fig. 3. Example of a SMF in the $\xi\eta$ plane at time t_1. For analyzer systems with ideal delay filters (see Fig. 2), the SMF is a δ point at position $\boldsymbol{\rho}_0$. For systems comprising low-pass filters, the SMF is a modulated δ line $u(\boldsymbol{\rho})$ with its center of gravity assumed at location $\boldsymbol{\rho}_c$.

$$U := \iint u(\boldsymbol{\rho}, t)\, \mathrm{d}\boldsymbol{\rho} = \int h(t)\, \mathrm{d}t =: H, \tag{11}$$

i.e., it is independent of time and thus is a constant that typifies the analyzer system.

The meaning of Eqs. (2c) and (2d) is now exemplified by an analyzer system with ideal delay filters of the impulse response $h_0(t) = \delta(t - \tau_0)$ [cf. formula (5)]. The SMF for such a system becomes

$$u_0(\boldsymbol{\rho}, t) = \delta[\boldsymbol{\rho} - \boldsymbol{\rho}_0(t)] := \delta[\xi - \xi_0(t)]\, \delta[\eta - \eta_0(t)], \tag{12}$$

where

$$\boldsymbol{\rho}_0(t) = \mathbf{R}(t) - \mathbf{R}(t - \tau_0) = [\xi_0(t), \eta_0(t)]^{\mathrm{T}} \tag{13}$$

and which is illustrated by the one-dimensional example in Fig. 2. When these results are applied to Eq. (2d), the MACF turns out to consist of adequately shifted ACF's:

$$k_0(\boldsymbol{\rho}, t) = k_a(\boldsymbol{\rho}) \ast^{\boldsymbol{\rho}} \delta[\boldsymbol{\rho} - \boldsymbol{\rho}_0(t)] = k_a[\boldsymbol{\rho} - \boldsymbol{\rho}_0(t)]. \tag{14}$$

Hence the SMF $u_0(\boldsymbol{\rho}, t)$ is a single δ-point function with the coordinates $\boldsymbol{\rho}_0(t)$ in the $\xi\eta$ plane (see Fig. 3). For nonperiodic object functions, the absolute maximum of the static ACF is always at its origin. Therefore, and owing to Eq. (14), the absolute maxima of the MACF in the $\xi\eta$ plane appear at the coordinates $\boldsymbol{\rho}_0(t)$. In other words, the shift vectors $\boldsymbol{\rho}_0(t)$ are determined exclusively by the SMF, and thus velocity estimation does not depend on the object function, a property that was pointed out in Section 3.

The insights gained in this section, especially the formulations of Eqs. (2c) and (2d), permit the derivation of a method for the velocity estimation in analyzer systems comprising realizable filters.

5. VELOCITY ESTIMATION IN SYSTEMS COMPRISING REALIZABLE TEMPORAL FILTERS

At every instant, SMF's of analyzer systems with band-limited filters are modulated δ-line functions in the $\xi\eta$ plane. They always start from the origin of the coordinate system and pass through point $\boldsymbol{\rho}_0$ (see Fig. 3). The curves depend on the combination of impulse response and true velocity. A single maximum of the SMF and the MACF in the $\xi\eta$ plane is now no longer guaranteed. Even if a mathematically unique maximum occurs at each moment, one confronts severe practical problems in attempting to determine its exact position in the MACF.[26] Therefore it is proposed to determine coordinates of comparable relevance through integral operations, namely, the centers of gravity (centroids) $\boldsymbol{\rho}_c(t)$ of the MACF (they should not be confused with the centroid of the object). As is known, the ACF of any real-valued object function is point symmetric and thus balanced with respect to the origin, i.e., its centroid is at the origin:

$$\iint k_a(\boldsymbol{\rho})\, \xi\, \mathrm{d}\boldsymbol{\rho} = \iint k_a(\boldsymbol{\rho})\, \eta\, \mathrm{d}\boldsymbol{\rho} \equiv 0 \qquad \text{for real-valued } b(\mathbf{r}). \tag{15}$$

Consequently, the centroid locations of the SMF are identical with those of the MACF, i.e., the centroid vectors $\boldsymbol{\rho}_c(t) = [\xi_c(t), \eta_c(t)]^{\mathrm{T}}$ are completely independent of the object form. Their vector components are defined by the integrals

$$\xi_c(t) = \tfrac{1}{K} \iint k(\boldsymbol{\rho}, t)\, \xi\, \mathrm{d}\boldsymbol{\rho} = \tfrac{1}{U} \iint u(\boldsymbol{\rho}, t)\, \xi\, \mathrm{d}\boldsymbol{\rho}, \tag{16a}$$

$$\eta_c(t) = \tfrac{1}{K} \iint k(\boldsymbol{\rho}, t)\, \eta\, \mathrm{d}\boldsymbol{\rho} = \tfrac{1}{U} \iint u(\boldsymbol{\rho}, t)\, \eta\, \mathrm{d}\boldsymbol{\rho}, \tag{16b}$$

with the constant factor

$$K = \iint k(\boldsymbol{\rho}, t)\, \mathrm{d}\boldsymbol{\rho} = U \iint k_a(\boldsymbol{\rho})\, \mathrm{d}\boldsymbol{\rho} = U\Big[\iint b(\mathbf{r})\, \mathrm{d}\mathbf{r}\Big]^2 =: UB^2. \tag{17}$$

The centroid of the filter impulse response,

$$\tau_c = \tfrac{1}{H} \int h(t) t\, \mathrm{d}t, \tag{18}$$

or any other characteristic time constant, and relation (11) permit the definition of a centroid-based velocity estimate

$$\mathbf{v}_c(t) = \boldsymbol{\rho}_c(t)/\tau_c\,. \tag{19}$$

It is a trivial fact that the position of a δ-point function coincides with its centroid location, and thus analyzer systems comprising ideal delay filters of impulse response $h_0(t) = \delta(t - \tau_0)$ represent special cases of the proposed approach. According to the relations (11) and (18) one obtains

$$\tau_0 = \int \delta(t - \tau_0) t\, \mathrm{d}t, \tag{20}$$

and finally, with the relations (11), (12), (2d), and (16), the components of the shift vector $\boldsymbol{\rho}_0(t)$ are given by

$$\xi_0(t) = \int \delta[\xi - \xi_0(t)]\, \xi\, \mathrm{d}\xi, \tag{21a}$$

$$\eta_0(t) = \int \delta[\eta - \eta_0(t)]\, \eta\, \mathrm{d}\eta. \tag{21b}$$

6. ACCURACY OF THE VELOCITY ESTIMATES

It may be useful to know about the accuracy of the velocity estimates that are delivered by a centroid-based analyzer system. Owing to the independence of the estimation process from object form, it is sufficient to investigate the evaluation of centroids from the SMF and not from the MACF.

In this section it is explained how well a temporally band-limited analyzer system approximates the temporal derivative of the trajectory

$$\mathrm{d}\mathbf{R}(t)/\mathrm{d}t = \mathbf{v}(t) = [v_x(t), v_y(t)]^{\mathrm{T}} = [X(t) * \delta'(t), Y(t) * \delta'(t)]^{\mathrm{T}}, \tag{22}$$

with

$$\delta'(t) = \mathrm{d}\delta(t)/\mathrm{d}t. \tag{23}$$

For this purpose it is feasible and convenient to consider the components of the trajectory $\mathbf{R}(t)$ separately. The following deductions are carried out for its x component $X(t)$, but they apply to the y component by analogy. The application of Eq. (16a) to the projection of the SMF onto the ξ axis,

$$u_\xi(\xi, t) = \int u(\boldsymbol{\rho}, t)\, \mathrm{d}\eta, \tag{24}$$

and taking into account relation (11), results in

$$H\,\xi_c(t) = \int u_\xi(\xi, t)\,\xi\, \mathrm{d}\xi. \tag{25}$$

As an example, the nonlinear transformation of a rectangular impulse response $h_r(\tau)$ with respect to the one-dimensional trajectory $X(t)$ is depicted in Fig. 4 for time $t = t_2$. The time-varying nonlinear transformation is mathematically formulated by

$$u_\xi[\xi(\tau), t] = h(\tau)/v(t - \tau). \tag{26}$$

Inserting this expression and the shift variable (cf. Fig. 2a)

$$\xi(\tau, t) = X(t) - X(t - \tau) \tag{27}$$

into Eq. (25), as well as integrating with respect to the new variable τ, yields

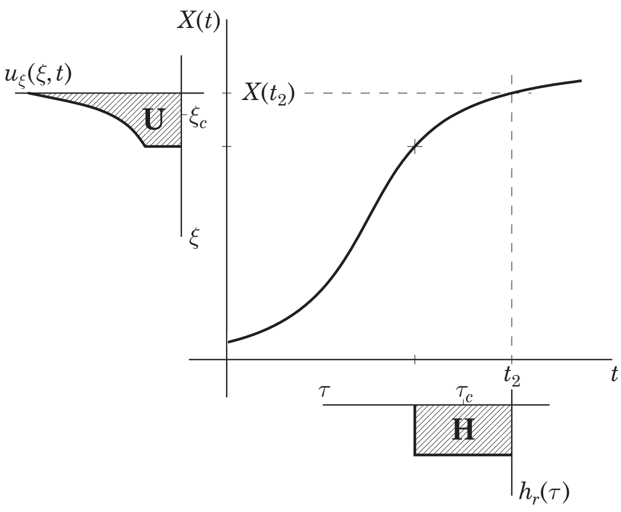

Fig. 4. Nonlinear transformation of the rectangular impulse response $h_r(\tau)$ with respect to the one-dimensional trajectory (see Fig. 2a) for time t_2.

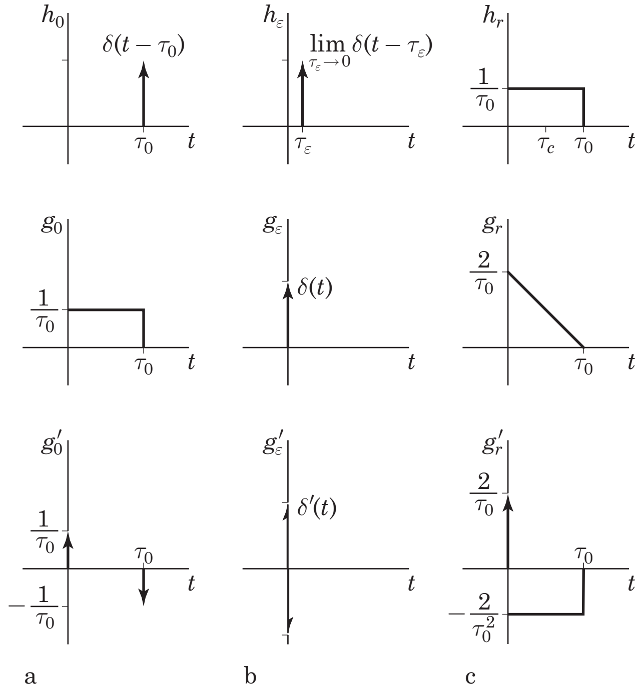

Fig. 5. Impulse responses $h(t)$ of three idealized filters (top), the corresponding kernel functions $g(t)$ (middle), and their derivatives $g'(t)$ (bottom). The estimated time course of the object velocity is the convolution of the true velocity profile with function $g(t)$. a, Ideal delay filter. b, Filter with an infinitesimal delay. c, Low-pass filter of rectangular impulse response.

$$\begin{aligned} H\,\xi_c(t) &= \int h(\tau)[X(t) - X(t - \tau)]\, \mathrm{d}\tau \\ &= X(t)\,H - \int h(\tau)\,X(t - \tau)\, \mathrm{d}\tau. \end{aligned} \tag{28}$$

Hence the ξ coordinate of the centroid location results from the convolution

$$\xi_c(t) = X(t) * \big[\delta(t) - h(t)/H\big]. \tag{29}$$

All convolutions in this section are purely temporal.

According to Eq. (19) the x component of the estimated velocity $\mathbf{v}_c(t) = [v_{cx}(t), v_{cy}(t)]^{\mathrm{T}}$ can be written as

$$v_{cx}(t) = X(t) * \tfrac{1}{\tau_c}\big[\delta(t) - h(t)/H\big] =: X(t) * g'(t), \tag{30}$$

or, with $g'(t) = \delta'(t) * g(t)$ and the relations (22) and (23),

$$v_{cx}(t) = X(t) * \delta'(t) * g(t) = v_x(t) * g(t). \tag{31}$$

By Eq. (31), the estimated velocity components $v_{cx,y}(t)$ are proved to result from the convolution of the true velocity components $v_{x,y}(t)$ with the time-invariant kernel

$$g(t) = \int_0^t g'(t)\, \mathrm{d}t = \tfrac{1}{\tau_c}\left[1\Big|_{t\geq 0} - \tfrac{1}{H}\int_0^t h(t)\, \mathrm{d}t\right], \tag{32}$$

which in turn depends on the integral function of the impulse response and thus is characteristic of the analyzer system. In short,

$$\mathbf{v}_c(t) = [v_x(t) * g(t), v_y(t) * g(t)]^{\mathrm{T}}. \tag{33}$$

The kernel functions $g(t)$ and their derivatives $g'(t)$ of three idealized analyzer systems are sketched in Fig. 5. A system with ideal delay filters leads to the rectangular kernel

$$g_0(t) = \tfrac{1}{\tau_0}\left[1\Big|_{t\geq 0} - \int_0^t \delta(t-\tau_0)\,\mathrm{d}t\right]$$
$$= \begin{cases} 1/\tau_0 & \text{for } 0 < t < \tau_0 \\ 1/(2\tau_0) & \text{for } t=0 \text{ and } t=\tau_0 \\ 0 & \text{otherwise} \end{cases} \tag{34}$$

and its corresponding derivative

$$g_0'(t) = \tfrac{1}{\tau_0}[\delta(t) - \delta(t-\tau_0)], \tag{35}$$

which are both shown in Fig. 5a. Figure 5b presents the kernel function for a system with infinitesimal delay,

$$g_\varepsilon(t) = \lim_{\tau_\varepsilon \to 0}\left\{\tfrac{1}{\tau_\varepsilon}\left[1\Big|_{t\geq 0} - \int_0^t \delta(t-\tau_\varepsilon)\,\mathrm{d}t\right]\right\} = \delta(t), \tag{36}$$

and its derivative,

$$g_\varepsilon'(t) = \delta'(t). \tag{37}$$

Thereby the measurement of the true velocity is confirmed [cf. formula (4)]. As a final example, a system with filters of rectangular impulse response $h_r(t) = g_0(t)$ with $\tau_c = \tau_0/2$ is considered. Such a system has the kernel functions depicted in Fig. 5c.

7. LIMITS AND MODIFICATIONS OF THE ANALYZER PRINCIPLE

This section deals with theoretical and practical limits of the analyzer principle and with some modifications necessary to allow for the analysis of more-realistic stimuli and for the explanation of visual phenomena, such as flicker fusion and apparent motion.

Owing to the normalization in Eqs. (16) and (18), centroid locations cannot be computed if the impulse response of the filter is of zero mean, $H = 0$ [cf. relations (11) and (17)]. Therefore the method fails for all kinds of high-pass or band-pass filters that have transfer functions with a zero at the origin. With respect to neural systems, the demand for $H \neq 0$ is not restrictive, since the direct representation of bipolar signals as impulse rates is impossible anyway. Besides, low-pass versions of delay filters seem more suited for this purpose and are biologically plausible.

The centroid-based approach permits, at least theoretically, the estimation of object velocity, independent of the object form. This holds even for motion of periodic object functions, as it may appear as background shifts caused by eye movements. In practice, however, broadband object spectra are advantageous. The signal energy of the ACF is then concentrated near its origin. Consequently, a greater resemblance between the SMF and the MACF is achieved, and the centroid extraction is facilitated, especially in systems of restricted computational accuracy. It is helpful as well to reduce the mean value of the input signals, as this is performed by the preprocessing stages of most biological systems. On the other hand, ideally differentiated object functions, i.e., those of zero mean, $B = 0$, again prohibit the evaluation of centroids [cf. relation (17)]. In neural systems, however, signal representations of zero mean are quite improbable.

The correlation analysis discussed so far can also be performed by local correlations, provided that it is restricted to speeds of less than a maximum value $v_{\max}$. For this purpose a minimum condition must be met: If for instance, the local area of analysis is assumed to be of the smallest possible extent 2ρ, then the centers of neighboring areas must not be separated by more than the distance ρ.[23,24] For an analyzer system characterized by τ_c, the maximum value $\rho_{\max}$ is defined by

$$\rho_{\max} = v_{\max}\,\tau_c\,. \tag{38}$$

There is indeed evidence for a limited range of distances ρ in the human visual system, for instance, Braddick's findings, which led to the distinction of short-range and long-range processes in motion perception.[27] For the motion analysis of more than a single object and of objects moving in front of structured backgrounds such locally computed MACF's are essential. They can simply be integrated in order to produce regional or global MACF's. In biological vision systems, this pooling is suspected to be object-specific.

The minimum detectable speed $v_{\min}$ in ρ-tuned analyzer systems with spatially discrete front ends is proportional to the resolution Δr of the object representation:

$$v_{\min} \propto \Delta r/\tau_c\,. \tag{39}$$

Van de Grind *et al.* actually found an eccentricity-dependent critical speed, which marks the lower speed limit v_{crit} for ρ-tuned detector mechanisms.[17] These investigations further reveal that, for lower speeds, the corresponding detector units are tuned by changing the time constant of the filter. It is therefore conjectured that $v_{\text{crit}} \approx v_{\min}$.

If an object stops moving, a δ-point function builds up at the origin of the $\xi\eta$ plane as part of the SMF. Sudden jumps of the object function, with static intervals, lead to SMF's that consist of a series of δ points with different impulse integrals. In the $v(t)$ diagram these ideal spatial jumps show up as δ functions, which, after their convolution with the kernel function $g(t)$, result in the time course of the velocity estimate. Under certain conditions human observers interpret such stimuli as continuously moving (apparent motion), although it is doubtful whether this phenomenon is explained merely by the smoothing effect of the kernel function. Objects remaining static for longer periods than the maximum duration $t_{\max}$ of the impulse response $h(t)$ lead to the static SMF,

$$u_s(\boldsymbol{\rho}, t) = u_s(\boldsymbol{\rho})\,\delta(t), \qquad u_s(\boldsymbol{\rho}) = H\delta(\boldsymbol{\rho}) \tag{40}$$

[see relation (11)], and, by using Eq. (2d), they are described by the static MACF,

$$k_s(\boldsymbol{\rho}, t) = k_s(\boldsymbol{\rho})\,\delta(t), \qquad k_s(\boldsymbol{\rho}) = Hk_a(\boldsymbol{\rho}). \tag{41}$$

In this case the analyzer system signals the correct centroid location $\boldsymbol{\rho}_c = \mathbf{0}$ and permits form analyses of the object based on its ACF, e.g., for the extraction of invariant features. When applied to other types of motion, such as expansions or contractions and rotations, correlative motion analysis results in the generalized MACF [cf. Eq. (44) below] and, for nonmoving objects, in the generalized ACF.[28-30] From the latter function, shape descriptors can easily be extracted, which are invariant under the whole group of similarity transformations and express object properties such as symmetries, similarities and congruences, i.e., categories known from Gestalt psychology.[23,29,31]

Intensity fluctuations of a moving object caused, e.g., by stroboscopic illumination can be described by a suitable function $f(t)$ that temporally modulates the δ-line function defined in Eq. (8). [Relation (11) does not hold for $f(t) \neq 1$.] For pure flicker, without any motion, the flicker SMF becomes

$$u_f(\boldsymbol{\rho}, t) = u_f(t)\,\delta(\boldsymbol{\rho}), \qquad u_f(t) = f(t)[f(t) \ast h(t)], \tag{42}$$

and thus the flicker MACF can be written as

$$k_f(\boldsymbol{\rho}, t) = u_f(t)\,k_a(\boldsymbol{\rho}). \tag{43}$$

As is known, the sensitivity of biological vision systems to flickered stimuli decreases with increasing flicker frequency. Therefore, and because Eq. (42) does not imply a band-limited flicker SMF, temporal low-pass filters must be assumed to exist in addition to the filter, preferably for both branches of each detector unit. This issue was discussed in detail by Foster.[32] Several detector models contain these filters (see, for instance, Fig. 1c of Ref. 2), but temporal filtering of the MACF, or combinations thereof, must also be considered. These filters prevent a sudden breakdown of the motion percept during short periods while $f(t) = 0$, for instance, in the case of momentary object occlusions.

8. SUMMARY AND CONCLUSIONS

A method is introduced that allows one to estimate the instantaneous velocity of an object that translates in a frontoparallel plane. For the mathematical analysis of the processing principle a spatially continuous array of isotropically arranged bilocal detector units is considered. These units comprise identical temporal filters and are tuned to different velocities through the span of their point sensors. At a first processing stage the output signals of the detector units are separately integrated according to their tuning velocities, leading to a signal representation that is described by the MACF. The analysis of other types of motion requires differently organized arrays of detector units; in the case of expansions, radial configurations, and for frontoparallel rotations, circular schemes are needed.[23,24] A general mathematical formulation is given by the generalized MACF of order 2:

$$k(\mathbf{z}, t) = \iint b[\mathcal{T}(\mathbf{z})\{\mathbf{r}\}, t]\, b_h(\mathbf{r}, t)\, \mathrm{d}\mathbf{r} \qquad \text{for real-valued } b(\mathbf{r}, t) \text{ and } b_h(\mathbf{r}, t). \tag{44}$$

The operator $\mathcal{T}(\mathbf{z})$ denotes geometric transformations, i.e., mappings $\mathbf{r} \to \mathbf{r}'$, that are characterized by the parameter vector $\mathbf{z}$. For the analysis of frontoparallel translations with $\mathbf{z} = \boldsymbol{\rho}$ one obtains $\mathcal{T}(\boldsymbol{\rho})\{\mathbf{r}\} = \mathbf{r} + \boldsymbol{\rho}$ and thereby the expression of Eq. (2a).

A second processing stage serves for the computation of the velocity estimate from the MACF; the position vector $\boldsymbol{\rho}(t)$ of either its maximum or its centroid is determined. Under certain conditions, namely, if periodic object functions are processed and for particular combinations of the velocity profile and the impulse response of the filters, maximum detection may lead to ambiguous results. The centroid-based scheme avoids these problems. In both cases the velocity estimate is the ratio of the position vector $\boldsymbol{\rho}(t)$ and a time constant τ that typifies the detector units. An important property of the estimation process is its independence from the form of the moving object—with slight restrictions in case of the maximum detection scheme.

An essential concern of the investigation is that of the clarification of the general interrelationship between the true velocity and its estimate. The latter turns out to be a low-pass version of the true velocity profile. The low-pass characteristic is a well-defined function of the impulse response of the filter and thus represents a time- and space-invariant property of the analyzer system.

As mentioned above, the geometric structure of the analyzer systems is specific for each type of motion. In conjunction with the correlative evaluation scheme, each analyzer system is selective for a certain kind of motion, i.e., it performs motion matching.[23,24] Thus certain kinds of compound motion can be decomposed according to the implemented analyzer systems (motion decomposition); each system signals the appropriate portion of the motion process with respect to its specific type of motion.[23] For instance, decomposition is not possible for frontoparallel rotations of arbitrary object functions that change in size.

This selectivity of correlative analyzer systems has consequences for the analysis of nonrigid objects as well. If, for example, a system for the analysis of frontoparallel translations is confronted with translations of a deforming object, it signals adequate deviations from the trajectory that would have been obtained from the corresponding rigid translations. There is no other choice because any changes in the detector plane are interpreted strictly in terms of translatory motion. Therefore it is feasible to extract aspects of arbitrary object motion in three-dimensional space by analysis of its planar projections. For a mathematical description of nonrigid object motion the object function is itself considered time dependent, i.e., the object function $b(\mathbf{r})$ in Eq. (2) is replaced by $b(\mathbf{r}, t)$. Now the MACF is obtained by the space-variant convolution

$$k(\boldsymbol{\rho}, t) = k_c[\boldsymbol{\rho}, \boldsymbol{\rho}'(\tau), t] \ast^{\boldsymbol{\rho}} u(\boldsymbol{\rho}, t), \tag{45}$$

where the shift vector

$$\boldsymbol{\rho}'(\tau, t) = \mathbf{R}(t) - \mathbf{R}(t - \tau) \tag{46}$$

[cf. Eq. (27) and Fig. 2] indicates the temporal variance of the set of cross-correlation functions

$$\begin{aligned} k_c(\boldsymbol{\rho}, \tau, t) &= \iint b(\mathbf{r} + \boldsymbol{\rho}, t)\, b_{pv}(\mathbf{r}, t - \tau)\, \mathrm{d}\mathbf{r} \\ &= b(\mathbf{r}, t) \otimes^{\boldsymbol{\rho}} b_{pv}(\mathbf{r}, t - \tau). \end{aligned} \tag{47}$$

At every time, function $k_c(\boldsymbol{\rho}, \tau, t)$ comprises the spatial cross-correlation functions of the present object function $b(\mathbf{r}, t)$ with all past views $b_{pv}(\mathbf{r}, t - \tau)$ of the object.

The correlative approach to motion analysis presented here is essentially based on the spatial integration of motion information. Because of this property it is suited to explaining well-known phenomena of biological vision, such as motion coherence and induced motion, in a way comparable with the approaches of Yuille and Grzywacz[33] as well as of Bülthoff *et al.*[5] Finally, it must be pointed out that, in biological systems, form independence of visual motion analysis can be regarded as a direct consequence of the crucial role that the common properties of real-world motion and movement stimuli play in the process of neural self-organization: Only great numbers of identical stimulations, i.e., repeated signal coincidences, lead to the consolidation of synaptic interconnections.

APPENDIX A: CONVOLUTIONS INVOLVING δ-LINE FUNCTIONS

The values F resulting from integrations over cross sections through δ-line functions play an essential role for convolutions involving such functions. In what follows it is shown that the parameter F_{xy}, which is relevant for the convolution of Eq. (9), is not a function of time t. This property is due to the specific formulation given by Eq. (8), which ensures that temporal fluctuations of the object intensity do not occur during object motion. According to Bamler,[34] the reciprocal integral value in a plane perpendicular to the tangent of the δ line equals the modulus of the outer product $\mathbf{w}(\mathbf{r}, t)$ of the three-dimensional normal vectors $\nabla a_x(\mathbf{r}, t)$ and $\nabla a_y(\mathbf{r}, t)$:

$$\mathbf{w}(\mathbf{r}, t) = \nabla a_x(\mathbf{r}, t) \times \nabla a_y(\mathbf{r}, t) = (w_x, w_y, w_t)^{\mathrm{T}}. \qquad \text{(A1)}$$

The normal vectors of the argument functions of Eq. (7) become

$$\nabla a_x(\mathbf{r}, t) = [1,0,-\partial X(t)/\partial t]^{\mathrm{T}} = [1,0,-v_x(t)]^{\mathrm{T}} \qquad \text{(A2a)}$$

as well as

$$\nabla a_y(\mathbf{r}, t) = [0,1,-\partial Y(t)/\partial t]^{\mathrm{T}} = [0,1,-v_y(t)]^{\mathrm{T}}. \qquad \text{(A2b)}$$

Their outer product turns out to be a vector that is independent of the xy coordinates:

$$\mathbf{w}(\mathbf{r}, t) = \mathbf{w}(t) = [v_x(t), v_y(t), 1]^{\mathrm{T}}. \qquad \text{(A3)}$$

The reciprocal modulus of its component w_t, i.e., the considered integral value

$$F_{xy} = 1/|w_t| = 1, \qquad \text{(A4)}$$

is indeed constant.

In order to compute the function $b_h(\mathbf{r}, t)$ one needs to know the integral value F_t, which depends on the components w_x and w_y of the vector $\mathbf{w}(t)$ [cf. Eq. (A3)]:

$$F_t = 1\Big/\sqrt{w_x^2 + w_y^2} = 1/v(t), \qquad v(t) = |\mathbf{v}(t)| > 0. \qquad \text{(A5)}$$

ACKNOWLEDGMENTS

Many helpful discussions with R. Bamler and H. Platzer are acknowledged. I thank E. Pöppel for his friendly support.

REFERENCES

1. J. P. H. van Santen and G. Sperling, "Temporal covariance model of human motion perception," J. Opt. Soc. Am. A **1**, 451-473 (1984).
2. J. P. H. van Santen and G. Sperling, "Elaborated Reichardt detectors," J. Opt. Soc. Am. A **2**, 300-321 (1985).
3. H. E. Adelson and J. R. Bergen, "Spatiotemporal energy models for the perception of motion," J. Opt. Soc. Am. A **2**, 284-299 (1985).
4. M. Egelhaaf and A. Borst, "Transient and steady-state response properties of movement detectors," J. Opt. Soc. Am. A **6**, 116-127 (1989).
5. H. Bülthoff, J. Little, and T. Poggio, "A parallel algorithm for real-time computation of optical flow," Nature (London) **337**, 549-553 (1989).
6. B. Hassenstein and W. Reichardt, "Systemtheoretische Analyse der Zeit-, Reihenfolgen- und Vorzeichenauswertung bei der Bewegungsperzeption des Rüsselkäfers Chlorophanus," Z. Naturforschg. **11b**, 513-524 (1956).
7. W. Reichardt, "Autokorrelations-Auswertung als Funktionsprinzip des Zentralnervensystems," Z. Naturforschg. **12b**, 448-457 (1957).
8. W. Reichardt, "Autocorrelation, a principle for evaluation of sensory information by the central nervous system," in *Principles of Sensory Communication*, W. A. Rosenblith, ed. (Wiley, New York, 1961), pp. 303-317.
9. W. Reichardt and D. Varjú, "Übertragungseigenschaften im Auswertesystem für das Bewegungssehen," Z. Naturforschg. **14b**, 674-689 (1959).
10. B. Hassenstein, "Optokinetische Wirksamkeit bewegter periodischer Muster," Z. Naturforschg. **14b**, 659-674 (1959).
11. G. Fermi and W. Reichardt, "Optomotorische Reaktionen der Fliege Musca domestica," Kybernetik **2**, 15-28 (1963).
12. J. Thorson, "Small-signal analysis of a visual reflex in the locust," parts I and II, Kybernetik **3**, 4-66 (1966).
13. V. Torre and T. Poggio, "A synaptic mechanism possibly underlying directional selectivity to motion," Proc. R. Soc. London Ser. B **202**, 409-416 (1978).
14. W. Reichardt, "Processing of optical information by the visual system of the fly," Vision Res. **26**, 113-126 (1986).
15. M. Egelhaaf and W. Reichardt, "Dynamic response properties of movement detectors: theoretical analysis and electrophysiological investigation in the visual system of the fly," Biol. Cybern. **56**, 69-87 (1987).
16. A. J. van Doorn and J. J. Koenderink, "The structure of the human motion detection system," IEEE Trans. Syst. Man Cybern. **SMC-13**, 916-922 (1983).
17. W. A. van de Grind, J. J. Koenderink, and A. J. van Doorn, "The distribution of human motion detector properties in the monocular visual field," Vision Res. **26**, 797- 810 (1986).
18. K. Nakayama, "Biological image motion processing: a review," Vision Res. **25**, 625-660 (1985).
19. P. I. Ruff, J. P. Rauschecker, and G. Palm, "A model of direction-selective 'simple' cells in the visual cortex based on inhibition asymmetry," Biol. Cybern. **57**, 147- 157 (1987).
20. W. Reichardt, "Evaluation of optical motion information by movement detectors," J. Comp. Physiol. **A 161**: 533-547 (1987).
21. W. Reichardt and M. Egelhaaf, "Properties of individual movement detectors as derived from behavioural experiments on the visual system of the fly," Biol. Cybern. **58**, 287-294 (1988).
22. W. E. Reichardt and R. W. Schlögl, "A two dimensional field theory for motion computation," Biol. Cybern. **60**, 23-35 (1988).
23. H. Glünder, "Invariant description of pictorial patterns via auto-comparison functions," doctoral dissertation (in German) (Technische Universität München, München, 1988); abstracts in IEEE Acoust. Speech Signal Process. Magazine **6**, 15 (1989); Signal Proc. **16**, 186 (1989). An extended summary in English is also available from the author.
24. H. Glünder, "A dualistic view of motion and invariant shape analysis," in *From Pixels to Features*, J. C. Simon, ed. (Elsevier, New York, 1989), 323-332.
25. A. Papoulis, "Optical systems, singularity functions, complex Hankel transforms," J. Opt. Soc. Am. **57**, 207-213 (1967).
26. H. T. Wang, B. Mathur, and C. Koch, "I thought I saw it move: computing optical flow in the primate visual system," in *Neuroscience and Connectionist Theory*, M. A. Gluck and D. E. Rumelhart, eds. (Erlbaum, Hillsdale, N.J., 1990), pp. 237-265.
27. O. Braddick, "A short range process in apparent motion," Vision Res. **14**, 519-527 (1974).

28. H. Glünder, A. Gerhard, H. Platzer, and J. Hofer-Alfeis, "A geometrical-transformation-invariant pattern recognition concept incorporating elementary properties of neuronal circuits," in *Proceedings of the 7th International Conference on Pattern Recognition*, M. Wein, ed. (IEEE Computer Society, Washington, D.C., 1984), pp. 1376-1379.
29. H. Glünder, "Invariant description of pictorial patterns via generalized auto-correlation functions," in *ASST '87, 6. Aachener Symposium für Signaltheorie*, D. Meyer-Ebrecht, ed. (Springer-Verlag, Berlin, 1987), pp. 84-87.
30. H. W. Strube, "A generalization of correlation functions and the Wiener-Khinchin theorem," Signal Proc. **8**, 63-74 (1985).
31. H. Glünder, "Neural computation of inner geometric pattern relations," Biol. Cybern. **55**, 239-251 (1986).
32. D. H. Foster, "A model of the human visual system in its response to certain classes of moving stimuli," Kybernetik **8**, 69-84 (1971).
33 A. L. Yuille and N. M. Grzywacz, "A computational theory for the perception of coherent visual motion," Nature (London) **333**, 71-74 (1988).
34. R. Bamler, *Mehrdimensionale lineare Systeme*. (Springer-Verlag, Berlin, 1989).