Analog VLSI motion projects at Caltech

Jörg Kramer and Giacomo Indiveri

University of Zurich, Institute of Neuroinformatics
Gloriastrasse 32, 8006 Zurich, Switzerland

Christof Koch

California Institute of Technology, Computation and Neural Systems 139-74
Pasadena, CA 91125, U.S.A.

ABSTRACT

We present different compact analog VLSI motion sensors that compute the 1-D velocity of optical stimuli over a large range and are suitable for integration in focal plane arrays. They have been extensively tested and optimized for robust performance under varying light conditions. Since their output signals are only weakly dependent on contrast, they directly extract optical flow data from an image. Focal plane arrays of such sensors are particularly interesting for application in single-chip systems that perform navigation tasks for moving robots or vehicles, where light weight, low power consumption, and real-time processing are crucial. Several monolithic motion-processing systems based on such velocity sensors have been built and tested. We describe here three chips, designed for the determination of the focus of expansion, the estimation of the time to contact, and the detection of motion discontinuities respectively. The first two systems have been specifically designed for vehicle navigation tasks. The choice of this application domain allows to make a priori *assumptions about the optical flow field that simplifies the structure of the systems and improves their overall performance. The motion-discontinuity-detection system can be more generally used to segment images based on the velocities of its different domains with respect to the camera. It is particularly useful for background-foreground segregation in the case of* ego-*motion of an autonomous system in a static environment. Test results of the three systems are presented and their performance is evaluated.*

Keywords: motion sensors, velocity sensors, optical flow, analog VLSI, photoreceptors, machine vision, robot vision, smart vision chips, parallel image processing

1. INTRODUCTION

Various applications in automotive navigation, robotics, and remote sensing require devices for processing visual motion that are small, consume little power, and work in real time. Since motion-sensing algorithms have a large computational overhead, most real-time machine-vision applications require special-purpose parallel hardware for computing motion across the entire image. Parallel hardware implementations are particularly attractive if image acquisition and motion computation can be integrated on a single silicon chip. Such *smart-vision* chips can be used in stand-alone applications, or as front-ends in conventional machine-vision systems.

Since image irradiance is a continuous function of time, asynchronous circuit implementations are preferable to clocked implementations. The latter introduce temporal aliasing artifacts that can significantly compromise time-sensitive computations, such as those associated with motion.

Analog processing is more economic in terms of silicon area and power than digital processing of comparable complexity, and thus allows higher pixel densities. Its main drawback is its lack of precision. Even an ideal system, however, would be unable to process motion with high precision, because of noisy input data and because of fundamental computational problems associated with the estimation of a velocity field from its optical flow field calculated from the spatio-temporal irradiance distribution, as obtained from a passive optical system.[1,2] We will argue that to obtain the type of information that can be robustly extracted from optical flow fields the limited precision of analog VLSI systems may thus be sufficient.

Several analog VLSI motion sensors have been proposed in the past ten years.[3–12] None of these are simultaneously compact enough for implementation in focal plane arrays, robust to background-brightness level, resistant against light flicker, insensitive to stimulus contrast and selective to velocity over a wide, unambiguous range.

In the first part of this article, we describe three motion sensors based on different algorithms that encode velocity over more than two orders of magnitude, independently of the global irradiance level over at least two orders of magnitude and of contrast for intermediate and high contrast values, while still responding to stimuli with optical contrasts as low as 5%.[13–15] Since these sensors are also compact, we can use arrays of them in monolithic systems for real-time machine-vision applications, that compute robust features of the optical flow field by means of integration and thresholding. As examples, we present in the second part of the article implementations of systems for computing the focus of expansion, time to contact and motion discontinuities respectively.[16,17]

2. ANALOG VLSI VELOCITY SENSORS

We implemented several velocity-measuring algorithms using analog VLSI technology. The three methods presented here are based on the measurement of the time of travel of an image feature from one location to another on the focal plane array. Such methods lend themselves nicely to analog VLSI implementations that discretize space into pixels, but unlike digital algorithms operate continuously in time. The velocity v of an image feature is inversely proportional to its time of travel Δt across a fixed pixel spacing Δx and can be calculated as

$$v = \frac{\Delta x}{\Delta t}. \tag{1}$$

2.1. Temporal edge detector

Our velocity sensors can be interfaced with any type of feature detector that outputs a spike-like signal in response to a given feature in real time. Since the simplest image features that can be detected by local measurements are edges, the current implementations use an edge detector as a front end.

We built an edge detector that operates in the temporal rather than the spatial domain, *i.e.* that detects irradiance changes at a single location instead of comparing the irradiance at nearby locations through lateral interaction. For this reason, we call it *temporal edge detector*.[13] As compared to spatial edge detectors, it is faster, smaller, and much less susceptible to fixed pattern noise, which is a big concern in compact analog VLSI array structures.

In its present version, the temporal edge detector converts a rapid increase in local brightness into a current pulse. This corresponds to detection of a temporal dark-bright or so-called ON edge. Note that a current pulse may also be triggered if the illumination over the entire chip changes simultaneously (light flicker). The edge detector consists of an adaptive photoreceptor[18] and a circuit that amplifies, differentiates, thresholds, and rectifies the signal. The output voltage V_{ph} of the photoreceptor increases logarithmically with image irradiance E. Thus, an irradiance transient dE causes a voltage transient dV_{ph} that is only a function of the contrast dE/E. This property is highly desirable for the extraction of local features from an image, because the overall illumination of a scene is likely to change with time. A capacitive gain stage amplifies the transient step response with respect to the adapted DC response by a factor A_{ph}. After the step, dV_{ph} slowly adapts to the DC response. The transient change dV_{ph} of the photoreceptor output voltage to an irradiance step dE is given by

$$dV_{ph} = A_{ph} V_K \cdot \frac{dE}{E} \tag{2}$$

with $V_K = V_T/\kappa$, where V_T denotes the thermal voltage and κ the back-gate coefficient.

The output of the photoreceptor is fed into a circuit that transduces positive voltage excursions in the photoreceptor voltage, corresponding to ON edges, to a current pulse I_e, which during sufficiently large ON-edge transients is given by

$$I_e = A_{amp} C_{amp} \cdot \frac{dV_{ph}}{dt} \cdot \Theta\left(\frac{dV_{ph}}{dt}\right), \tag{3}$$

where A_{amp} denotes the amplifier gain, C_{amp} the capacitance of the differentiator circuit, and $\Theta(.)$ the Heaviside function. Substituting eq. (2) we obtain

$$I_e = A_{amp} A_{ph} C_{amp} V_K \cdot \frac{dE}{E dt} \cdot \Theta\left(\frac{dE}{dt}\right). \tag{4}$$

The term dE/Edt can be regarded as *temporal contrast*, which for a moving stimulus is the product of its velocity v and its spatial contrast dE/Edx, so that in a

motion sensor

$$I_e \sim \frac{dE}{Edt} = v \cdot \frac{dE}{Edx}\,. \tag{5}$$

In order to provide maximum sensitivity and at the same time a robust output signal the photoreceptor has to be optimized to the expected contrast and velocity range of the optical stimuli. Since the height of the output current pulse I_e of the edge detector is proportional to the slew rate dV_{ph}/dt of the photoreceptor output voltage V_{ph}, as evidenced by eq. (3), the slew rate needs to be maximized. On the other hand, overshoot or fast adaptation effects of V_{ph} cannot be tolerated, since this would cause spurious output pulses from the edge detector.

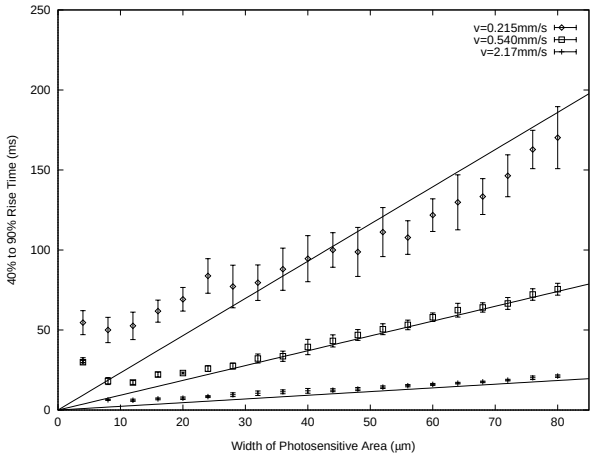


Fig. 1. Rise time of the photoreceptor output voltage in response to a moving edge stimulus as a function of photodiode size for different stimulus speeds. For small photodiode sizes the rise time is bandwidth-limited, for larger ones it is limited by the edge-transition time across the photodiode (solid lines).

To determine the optimum photodiode size we measured the bandwidth B and the rise time τ for photoreceptors with square photodiodes of different side lengths a, ranging from 4 μm to 80 μm. The bandwidth measurements were performed using a DC-driven LED to generate the optical stimulus. At typical illumination levels, the $1/f$ noise turned out to be insignificant and we obtained lowpass-filtered white-noise characteristics. The rise-time measurements were made using a square-wave modulated LED. Since the photocurrent is proportional to a^2E and the photodiode node capacitance is proportional to a^2+p, where p is an an area-independent contribution from the parasitics, we expect[19]

$$B \sim \tau^{-1} \sim \frac{a^2}{a^2+p}E\,. \tag{6}$$

Experimentally, we found

$$B \sim \tau^{-1} \sim a^2E\,, \tag{7}$$

for all photoreceptors but the ones with the largest photodiodes ($a \geq 60$ μm). The parasitics thus dominate the node capacitance for small and intermediate photodiode sizes ($p \gg a^2$). Bandwidth considerations therefore favor large photodiodes.

For a moving edge stimulus, however, the rise time of V_{ph} is also limited by the finite edge-transition time across the photodiode to $\tau \geq a/v$, where v denotes the velocity of the edge image on the chip. In order to investigate the combined effect of bandwidth and transition-time limitation, we projected a moving pattern of black and white stripes via a microscope onto the chip containing the different photoreceptors. Fig. 1 shows the measured rise times of the photoreceptor output signal as a function of the photosensitive area for three different velocities. For small areas, the $1/a^2$ bandwidth dependence dominates, whereas for larger areas (a $\geq$ 30 μm) the rise time becomes linear with a, approaching the transition-time limit, indicated by the solid lines. For the speed range between 0.2 mm/s and 25 mm/s the rise time was minimum for $a \approx 10-20$ μm. Assuming a pixel pitch of 60 μm that we typically used for our arrays, this speed range corresponds to a range of about 3 pixels/s to 400 pixels/s.

2.2. FS algorithm

The schematic architecture of one functional element of the FS motion sensor[13,14] is shown in Fig. 2(a). The output current pulses (I_1 resp. I_2) of two adjacent edge detectors (E) are fed into pulse-shaping circuits (P) that generate a slowly decaying voltage signal (V_{f1} resp. V_{f2}) and a sharp voltage spike (V_{s1} resp. V_{s2}) in response to each edge pulse. In a motion-sensing stage, the slowly decaying signal V_{f1} from one location is read upon the arrival of the sharp spike V_{s2} from the adjacent location and its value V_r is stored (and *vice versa* with V_{f2}, V_{s1}, and V_l). Since the slowly decaying signal can be regarded as a facilitation pulse for the velocity measurement, that is sampled by the spike signal, we call the algorithm *facilitate and sample (FS)*. The motion-sensing stage consists of two elements (M), each estimating velocity for one direction of motion of the edge. The element for which the onset of the facilitation pulse precedes the sampling spike (Fig. 2(b)) encodes the relative time delay of these signals generated by the same edge at the two locations and therefore, according to eq. (1),

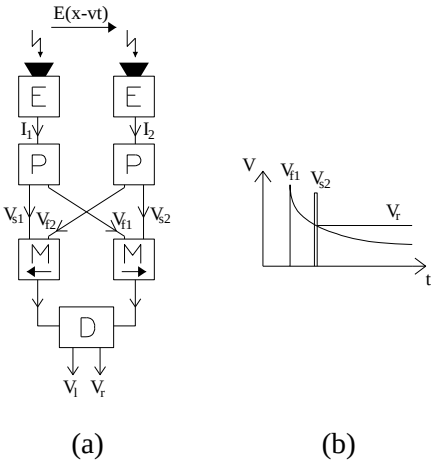


Fig. 2. Facilitate-and-sample (FS) motion sensor. (a) Block diagram. Temporal edge detectors (E) generate a current pulse in response to a fast image brightness transient. Pulse-shaping circuits (P) convert the current pulse into two voltage signals. Voltage signals from adjacent pixels are fed into two motion cells (M) computing velocity for opposite directions of motion along one dimension. A direction-selection circuit (D) suppresses the response in the null direction to prevent temporal aliasing. (b) Voltage signals. For a rightward-moving edge, the analog output voltage V_r of the right motion cell equals the voltage V_{f1} of the slowly decaying facilitation pulse initiated by the left edge detector at the time of arrival of the short sampling pulse V_{s2} initiated by the right edge detector. For leftward motion, V_{s2} precedes V_{f1} and V_r is low. V_r thus encodes velocity for rightward motion and V_l correspondingly for leftward motion, using V_{f2} as facilitation pulse and V_{s1} as sampling pulse.

the edge velocity. The edge is then said to move in the *preferred direction* of motion of the element. The monotonic decay of the facilitation pulse ensures unambiguous encoding of speed in the sampled voltage. For the other element, the sampling spike precedes the onset of the facilitation pulse and the residual voltage of the facilitation pulse triggered by the previous edge is sampled. The edge is said to move into the *null direction* of this element and the sampled voltage does not contain any velocity information. In absence of *spatial aliasing*, *i.e.* if the separation of succeeding edges is larger than twice the pixel spacing, the output voltage in the preferred direction is always greater than the spurious output voltage in the null direction, which can thus be suppressed by comparing the voltages at the two outputs and setting the lower one to zero, using a comparator circuit (D).

The pulse-shaping circuit (P) is implemented as the nonlinear filter shown in Fig. 3, consisting of a low-pass filter (Q_3–Q_5, C) in the feedback path of a high-gain amplifier (Q_1, Q_2). An input current pulse I_{in} (a mirrored and amplified copy of I_e of the edge detector) gives rise to a sharp spike V_s (sampling pulse) at the amplifier output and a pulse V_f (facilitation pulse) with a sharp onset and a slow decay at the low-pass filter output. For an input pulse amplitude of

$$I_o = (I_{in})_{max} = A_{mirr}(I_e)_{max} \,, \tag{8}$$

where A_{mirr} denotes the gain of the current mirror, we obtain

$$V_f(t) = V_{K2} \cdot \log\left(\frac{\frac{I_o}{I_n}}{1 + \frac{I_o t}{C V_{K2}}}\right) \,, \tag{9}$$

where $V_{K2} = V_T(\kappa + 1)/\kappa^2$ and where I_n is a process-dependent current-scaling parameter for Q_4 and Q_5. Note that after a time $t \gg CV_{K2}/I_o$, V_f is independent of I_o and therefore of temporal edge contrast (*cf.* eq. (4)). The decay of V_f does not have a time constant

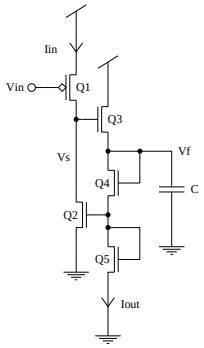


Fig. 3. Schematic diagram of the pulse-shaping circuit (P). It is a nonlinear filter generating a short voltage pulse V_s and a slowly decaying voltage pulse V_f in response to a current pulse I_{in} that is an amplified mirror copy of I_e of the temporal edge detector.

determined by a bias voltage, since a diode-capacitor configuration intrinsically changes the decay time constant over many orders of magnitude.

The motion cells (M) are standard sample-and-hold circuits where the facilitation pulse V_f from one location is read upon the arrival of the the sampling pulse V_s from the adjacent location and stored on a capacitor until the value gets updated by the next sampling pulse.

The null-suppression circuit (D) is a high-gain comparator that feeds the higher output voltage through a pass transistor and shorts the terminal with the lower one to ground.

The sampled output voltage V for the preferred direction (*i.e.* V_r resp. V_l) as a function of the velocity v in the focal plane can be computed from eqs. (1) and (9) as

$$V = V_{K2} \cdot \log\left(\frac{\frac{I_o}{I_n}}{1 + \frac{I_o/I_n}{v/v_o}}\right) \,, \tag{10}$$

where $v_o = \Delta x I_o / C V_{K2}$ can be regarded as a velocity-scaling parameter. As long as the photoreceptor is not bandwidth-limited we can substitute I_o using eqs. (4), (5), and (8) to obtain

$$V = V_{K2} \cdot \left[\log \left(\frac{x_o \frac{(dE)_{max}}{Edx}}{1 + x_o \frac{(dE)_{max}}{Edx}} \right) + \log \left(\frac{v}{v_o} \right) \right] , \quad (11)$$

where $x_o = \Delta x A_{mirr} A_{amp} A_{ph} C_{amp} V_K / C V_{K2}$ is a spatial scaling parameter. In eq. (11) the contrast dependence and the velocity dependence of the output voltage are separated into two additive terms. For spatial contrasts $\frac{(dE)_{max}}{Edx} \gg x_o^{-1}$ the output voltage V is independent of contrast and given by

$$V = V_{K2} \cdot \log \left(\frac{v}{v_o} \right) . \quad (12)$$

This output characteristic features a good sensitivity over a large range of velocities ($dV/dv \sim v^{-1}$). For typical process parameters, $x_o \approx 40\,\Delta x$ and the condition for contrast-independence is fulfilled for gradual edges with spatial contrasts much larger than 2.5% per pixel. For sharp edges, the measured spatial contrast is limited by the finite size of the photosensitive area of each pixel. Assuming the typical values $\Delta x = 60$ μm for the pixel spacing and $a = 24$ μm for the linear extension of the photosensitive area, that we used for linear motion-sensing arrays, the condition for contrast independence becomes $\frac{dE}{E} \gg 1\%$. For smaller contrasts, the response curve shifts toward smaller values of V.

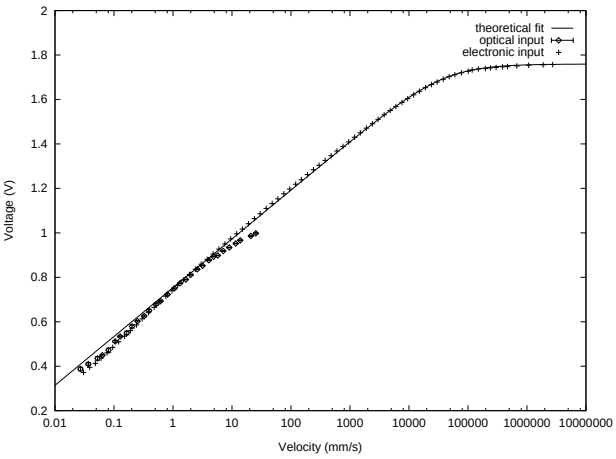


Fig. 4. Response of the FS motion sensor for optical and electronic stimuli. The output voltage for motion in the preferred direction as a function of image velocity for a 56% contrast, sharp ON edge under AC-incandescent room illumination is compared to that measured for stimulation of the motion-computing circuitry with electronically generated pulses to simulate the outputs of adjacent edge detectors to idealized edges. Each data point represents the average of five successive measurements. A theoretical fit to the latter data, based on eq. (10) is also shown. The r.m.s. error of the data points for optical input, as indicated by error bars, averages to about 3 mV, whereas for electronic stimulation each data point has an r.m.s. error of less than 1 mV. The FS motion sensor shows good performance over at least eight orders of magnitude of velocity for standardized input signals. A major improvement would thus be the design of an edge detector providing such contrast-independent pulses.

At low illumination levels, the photoreceptor starts to bandwidth-limit (*cf.* eq. (6)) and I_o becomes independent of v. Eq. (10) then reduces to $V = V_{K2} \cdot \log(I_o / I_n)$ for $v \gg v_o I_o / I_n$, *i.e.* for high velocities the response curve asymptotes to a constant output voltage.

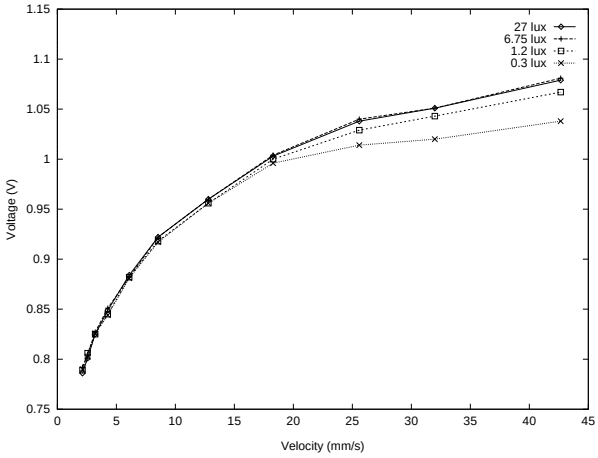


Fig. 5. Output voltage of the FS motion sensor for the preferred direction of motion of a 56% contrast, sharp ON edge versus image velocity for different AC-incandescent room-illumination levels. Each data point represents the average of five successive measurements. The saturation towards high speeds for low illumination levels is due to bandwidth limitation of the edge detector.

For quantitative measurements, sheets of paper with printed gray-scale patterns were wrapped around a rotating drum to provide the optical stimuli. The stimuli were projected onto the chip from a distance of 380 mm by means of a lens with a focal length $f = 13$ mm and an f-number of 1.8. Unless otherwise noted, measurements were taken under AC-incandescent room lighting conditions, where a white paper surface provided an illuminance of about 1.2 lux on the chip. The measurements were performed on a velocity sensor fabricated with a 2 μm n-well CMOS process with $a = 24$ μm and $\Delta x = 300$ μm. The gain of the temporal edge detector was reduced (by operating the amplifier circuit above threshold) to a value where it did not respond to the 120 Hz flicker noise of the light source.

Fig. 4 shows the response of the FS sensor to a 56%

contrast, sharp ON edge as a function of the velocity of its image on the chip in the preferred direction and its response to pairs of electronically generated binary pulses fed into the pulse-shaping circuits (at V_{in} in Fig. 3), simulating motion through their relative delay. A theoretical fit, based on eq. (10), is also plotted. The width of the input pulses was set to 50 μs. The velocity range for the optical stimulus was limited to about three orders of magnitude — at the low end by the temporal contrast of the stimulus falling below the threshold of the edge detector and at the high end by the maximum speed we could generate with the rotating drum. For the electronic stimulation, the output voltage is a monotonic function of simulated velocity over at least eight orders of magnitude with an r.m.s. noise of less than 1 mV. The output voltage is a logarithmic function of velocity, as given by eq. (12), except for small time delays, where the condition for contrast-independence is not satisfied and eq. (10) differs from eq. (12), and for large time delays, where the voltage decays because of parasitic leak currents to ground.

The response curves of the FS circuit to the ON edge at different global illumination levels are shown in Fig. 5. As expected, the curves for low illumination saturate towards high velocities, due to the bandwidth limitation of the photoreceptor. Fig. 6 shows the effect of using

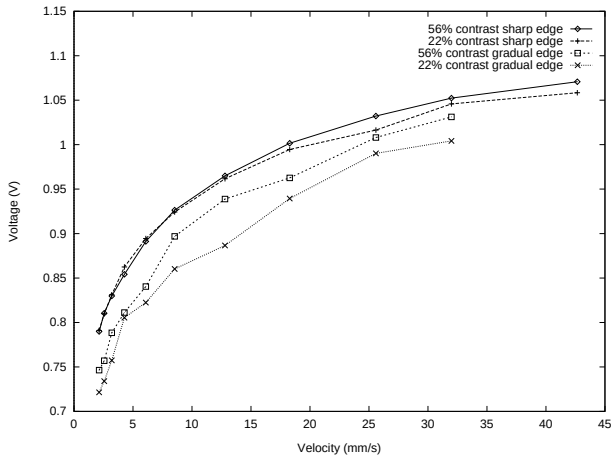


Fig. 6. Output voltage of the FS motion sensor for the preferred direction of motion of different ON edge stimuli versus image velocity for AC-incandescent room illumination. Each data point represents the average of five successive measurements. The data shows the contrast dependence of the velocity measurements predicted by eq. (11).

different edge sharpnesses and contrasts. The edge contrast is defined here as the ratio of the difference and the sum of maximum and minimum irradiance. For small local contrasts, the response curve shifts to lower output voltages, as predicted by eq. (11), and the sampling pulse becomes broader, introducing more jitter in the sampling time and thus more noise. The measured response of the FS circuit to sharp edges is reasonably contrast-invariant down to contrasts of approximately 15%, with a detection limit at about 5% contrast. If DC illumination is used, there is no 120 Hz flicker noise, allowing us to increase the gain of the temporal edge detector. This further lowers the limits for contrast invariance and detection.

2.3. FT algorithm

The schematic architecture of the FT motion sensor is shown in Fig. 7(a).[14] The motion-sensing principle is similar to that of the FS algorithm. Here, the pulse-shaping circuits (P) generate a binary voltage pulse of a given duration (V_{p1} resp. V_{p2}) using a neuron circuit. The algorithm is called *facilitate and trigger (FT)*, because for each motion cell (M) the pulse from one location acts as a facilitation pulse and the one from the other location as a trigger pulse, initiating at its onset an output pulse (V_r resp. V_l) if the facilitation pulse is high. The output pulse is terminated when the facilitation pulse goes low, as shown in Fig. 7(b). No output pulse is generated if the facilitation signal is low at the

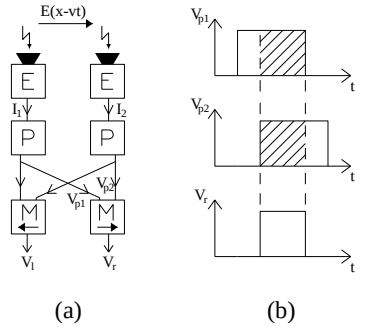


Fig. 7. Facilitate-and-trigger (FT) motion sensor. (a) Block diagram. Temporal edge detectors (E) generate a current pulse in response to a fast image brightness transient. Pulse-shaping circuits (P) convert the current pulse into a voltage pulse. The voltage pulses from adjacent pixels are fed into two motion cells (M) computing velocity for opposite directions of motion along one dimension. (b) Voltage pulses. For a rightward-moving edge, the width of the output pulse V_r of the right motion cell equals the overlap time of the facilitation pulse V_{p1} initiated by the left edge detector and the trigger pulse V_{p2} initiated by the right edge detector. For leftward motion, V_{p2} precedes V_{p1} and no output pulse is generated. V_r thus encodes velocity for rightward motion and V_l correspondingly for leftward motion, using V_{p2} as facilitation pulse and V_{p1} as trigger pulse.

onset of the trigger pulse. The preferred direction of motion for each cell is thus the direction for which the onset of the facilitation pulse precedes the onset of the trigger pulse. In the preferred direction, the duration τ of the output pulse equals the temporal overlap of the facilitation and the trigger pulse according to

$$\tau = \left(T - \frac{\Delta x}{v}\right) \cdot \Theta\left(T - \frac{\Delta x}{v}\right) , \qquad (13)$$

where T denotes the width of the neuron pulse. There is no output pulse if the travel time $\Delta x/v$ of the edge between the two edge detectors is longer than the facilitation pulse.

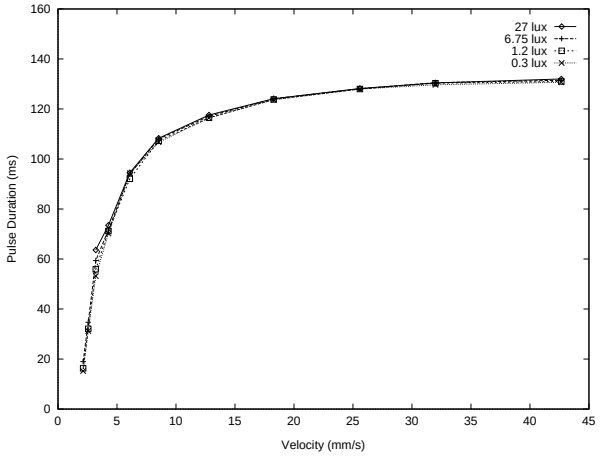


Fig. 8. Output pulse duration of the FT motion sensor for the preferred direction of motion of a 56% contrast, sharp ON edge versus image velocity for different AC-incandescent room-illumination levels. Each data point represents the average of five successive measurements. The FT sensor is less illumination-sensitive than the FS sensor.

The response characteristics of the FT sensor, as obtained with the experimental set-up described in Section 2.2., are shown in Fig. 8 for different global illumination levels and in Fig. 9 for different edges. We note that the FT sensor is less susceptible to illumination and contrast variations than the FS sensor, which is mainly due to the fact that the time of travel is now directly encoded in the time domain instead of being converted into an analog voltage. However, the sensitivity of the FT response curve decreases rapidly with increasing velocity, according to $d\tau/dv \sim d\tau/\tau dv \sim v^{-2}$, and is highest near the low-velocity cut-off $\Delta x/T$. For applications with smoothly varying velocity fields this problem can be alleviated by making the pulse duration T adapt to a value proportional to the average duration $\overline{\tau}$ of the preceding output pulses, a good choice being $2\overline{\tau}$. The adaptation does not change the absolute sensitivity $d\tau/dv$, but improves the relative sensitivity characteristic to $d\tau/\tau dv \sim v^{-1}$ and dynamically changes the low-velocity cut-off.

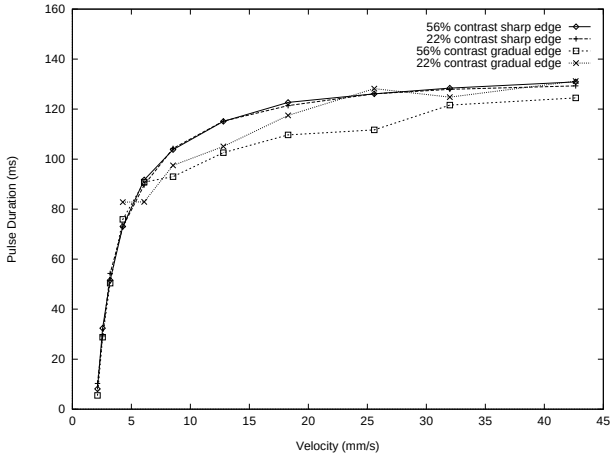


Fig. 9. Output pulse duration of the FT motion sensor for the preferred direction of motion of different ON edge stimuli versus image velocity for AC-incandescent room illumination. Each data point represents the average of five successive measurements. The FT sensor is less contrast-sensitive than the FS sensor.

2.4. FTI algorithm

The schematic architecture of the FTI motion sensor is shown in Fig 10.[15] Unlike the FS and FT algorithms

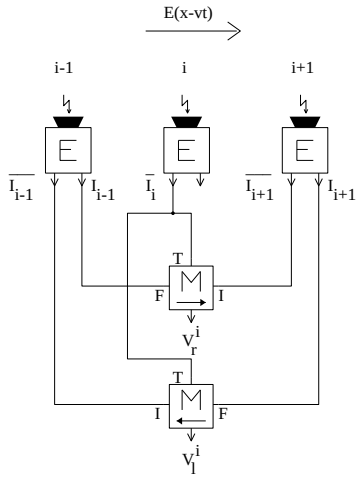


Fig. 10. Block diagram of the facilitate-trigger-and-inhibit (FTI) motion sensor. Current spikes from edge detectors (E) at three adjacent pixel locations $i-1$, i, $i+1$ are used as facilitation (F), trigger (T), and inhibition (I) signals for output pulses V_r^i and V_l^i of direction-selective motion cells (M).

the FTI algorithm directly uses the signals from three adjacent edge detectors for motion computation. An

edge moving from left to right causes the edge detector at position $i-1$ to facilitate the response of the circuit for rightward motion with a current spike I_{i-1}. Upon detection of the edge at position i, a current spike $\overline{I}_i$ is produced, that triggers a voltage pulse V_r^i at the output of the motion circuit. When the edge arrives at position $i+1$, the voltage pulse is terminated by inhibitory action induced by $\overline{I}_{i+1}$. The algorithm is thus called *facilitate trigger and inhibit (FTI)*. The output-pulse duration Δt is inversely proportional to the velocity v of the edge stimulus according to eq. (1). If the edge moves

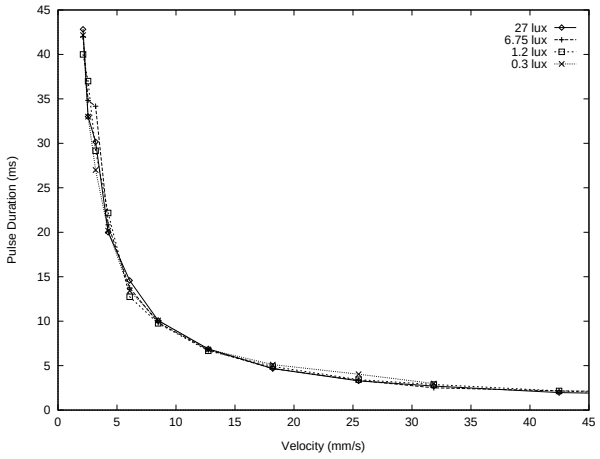


Fig. 11. Output pulse duration of the FTI motion sensor for the preferred direction of motion of a 56% contrast, sharp ON edge versus image velocity for different AC-incandescent room-illumination levels. Each data point represents the average of six successive measurements. Like the FT sensor, it is almost illumination-invariant in the measured range.

from right to left, the current spike $\overline{I}_{i+1}$ is generated first and inhibits the triggering of an output pulse V_r^i by the succeeding current spike $\overline{I}_i$. The inhibition is terminated by the facilitation spike I_{i-1} when the edge is detected at position $i-1$. The circuit for leftward motion has its facilitatory input connected to I_{i+1} and its inhibitory input to $\overline{I}_{i-1}$. It therefore only generates an output pulse V_l^i for edges moving from right to left.

The FTI circuit does not use any internal time constants for the measurement of velocity. The facilitation and inhibition time intervals are intrinsically adaptive over a large velocity range, since they are directly determined by the velocity of the edge, that triggers facilitation and inhibition upon its arrival at nearby locations on the chip, whereby inhibition terminates facilitation and *vice versa*. Furthermore, the fixed-pattern noise of an FTI sensor array is very low, because the velocity measurement does not depend on any circuit parameters that show a large variability across the array.

Measurements on an FTI sensor array were made with the experimental set-up used for the characterization of the FS and FT sensors, that is described in Section 2.2. The pixel pitch of the array was $\Delta x = 85\ \mu$m. The results for different illumination levels and edges are shown in Figs. 11 and 12 respectively. Like the FT sensor, the FTI sensor is very insensitive against changes of illumination and quite robust against contrast variations. Its sensitivity characteristics are the same as for the adaptive version of the FT sensor, but with reversed sign, *i.e.* the output pulse width decreases with increasing velocity. The FTI sensor is significantly more compact than the FS and the FT sensor and thus more suitable for dense focal plane arrays.

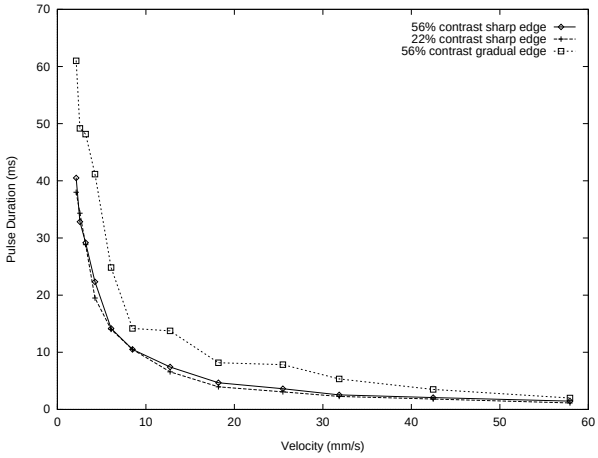


Fig. 12. Output pulse duration of the FTI motion sensor for the preferred direction of motion of different ON edge stimuli versus image velocity for AC-incandescent room illumination. Each data point represents the average of six successive measurements. The performance is similar to that of the FT sensor.

3. SYSTEM IMPLEMENTATIONS

A velocity field is of limited practical use in the absence of further processing, since it contains a large amount of redundant data, from which the useful information has first to be extracted. Furthermore, the optical flow field that can be measured using irradiance data obtained from passive optical sensors is at best an approximation of the projection of the velocity field onto the sensing surface, even under the assumption that the sensors and the processing system are ideal.[1,2]

In this section, we present three system implementa-

tions, that show how 1-D arrays of velocity sensors can be used to extract valuable information from the optical flow field in the focal plane. In all three systems we use *a priori* knowledge about the environment to extract from the large amount of inaccurate and redundant velocity information low-dimensional, more robust, and directly useful features of the velocity field, using averaging and thresholding operations.

Each of these systems employs an array of FS velocity sensors as an input stage. This algorithm was chosen because of its good sensitivity over a large velocity range and because its voltage representation that retains the last measured value lends itself more conveniently to parallel computation than the time-domain pulse representation of the FT and FTI algorithms.

3.1. Focus of Expansion

During observer motion through the environment, the velocity vectors generated in an instant of pure translational motion are radial in nature and expand out from a point that corresponds to the direction of heading, also referred to as the *focus of expansion (FOE).*[20] For vehicle-navigation applications we can restrict our analysis to pure translational motion taking advantage of the fact that it is possible to compensate for the *rotational* component of motion using inertia sensors often already present on the vehicle (*e.g.* for controlling the ABS system). Being interested in controlling the heading direction along the horizontal axis we can further reduce the complexity of the problem by considering 1-D arrays of velocity sensors. In such a case, where only the horizontal component of the optical flow vectors obtained from pure translational motion in a fixed environment is measured, the problem of detecting the FOE reduces to detecting the point in which the optical flow vectors change direction. Coding such vectors with positive values for one direction and negative values for the opposite direction, the problem then is to detect the *zero-crossing* in the data array. To analyze the computational properties of the optical flow for typical vehicle-navigation scenes in real cases, we performed software simulations on sequences of images obtained from a camera with a 64×64 pixel silicon retina placed on a moving truck (provided by Rockwell Corp.).[21] In the software simulations the optical flow-field vectors were obtained using a standard machine-vision method based on the *brightness-constancy equation*. The horizontal components of the vectors were encoded with positive/negative values depending on their direction and averaged along the image columns. To efficiently detect the zero-crossings corresponding to the correct FOE position we simulated an algorithm that uses the following operators: saturating non-linearity on the horizontal components of the optical flow vectors (only the sign of the vector is needed); spatial smoothing on the resulting values; evaluation of the steepness of the zero-crossings present in the array and finally selection of the zero-crossing with maximum steepness. The zero-crossing with maximum steepness will be selected only if its position is in the neighborhood of the previously selected zero-crossing. This ensures that the computed heading-direction position will shift smoothly with time.

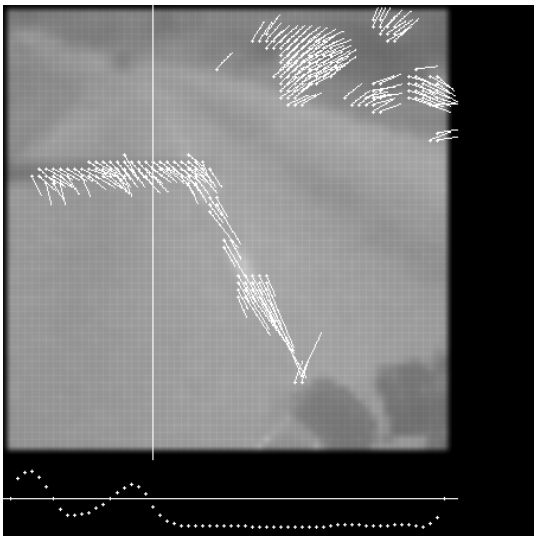


Fig. 13. Image obtained from a camera mounted on a truck, moving on a straight road, with optical flow vectors superimposed. The bottom of the figure represents the sum of the horizontal components of the optical flow field for each column. The coordinate of the heading direction is computed as the abscissa of the zero-crossing with maximum steepness and closest to the abscissa of the previously selected unit.

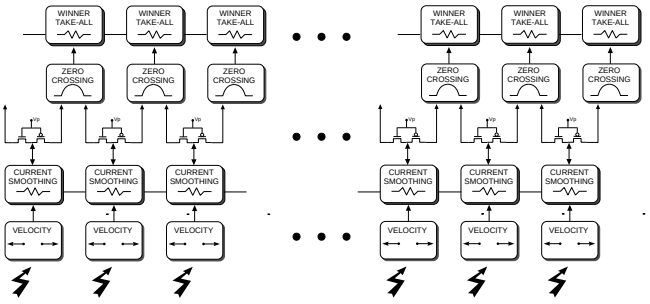
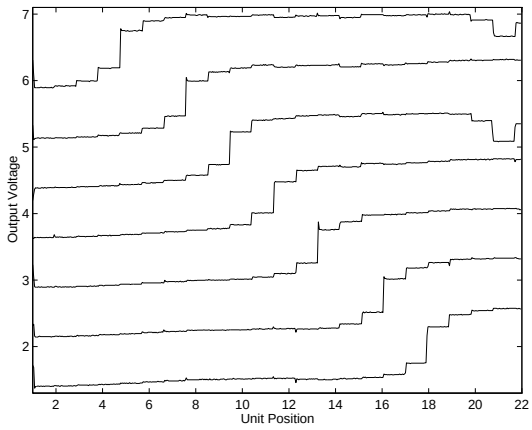
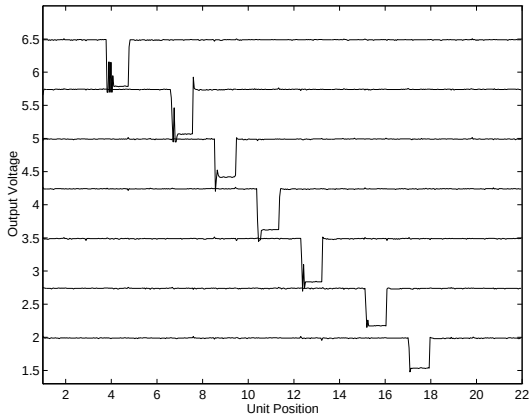


Fig. 14. Block diagram of the analog VLSI architecture for determining the FOE position for an observer translating in a fixed environment.

Fig. 13 shows a simulation result for an image taken from the straight road sequence. All of the operators used in the algorithm have been carefully chosen to be suitable for implementation with analog circuits. We implemented the algorithm on a 2.23 mm $\times$ 2.23 mm chip, using a 2 μm CMOS process. A block diagram of the



(a)



(b)

Fig. 15. (a) Output of the FOE system at the second processing stage (direction-of-velocity selection and differential-voltage-to-current conversion) as the FOE position shifts from left to right. The output currents were converted into voltages using an off-chip sense-amplifier with a reference voltage set to 2 V. Each plot was separated from the next by adding a constant value of 0.75 V. (b) Output of the system at the third processing stage (zero-crossing detection) for the same sequence of FOE positions.

hardware architecture that implements the algorithm described above is shown in Fig. 14. There are four main processing stages: the input stage is a 1-D array of FS velocity sensors that measure speed and direction of motion of temporal edges. The differential voltage signal of each velocity sensor is then converted (in the second processing stage) into a current, by means of a circuit based on a two-node winner-take-all (WTA) network.[22] Depending on the direction of motion of the stimulus, the current will be sourced either in the left branch of the circuit or in the right one. Spatial smoothing is also implemented in this stage using "diffusor" networks[23] that laterally spread the currents coming from the WTA branches. Finally, still in the second stage, the smoothed currents coming from the two branches are subtracted one from the other so that stimuli moving in one direction are encoded with positive currents and stimuli moving in the opposite direction are encoded with negative ones. Zero-crossings are detected in the third processing stage by looking for co-presence of negative currents from one unit and positive currents from the neighboring unit. The circuit that implements this operation is based on the current-correlator circuit [24]: if both input currents are greater than a set threshold value, an output current, corresponding to the sum of the absolute value of the two inputs, is generated. To select the zero-crossing corresponding to the correct FOE location we choose the one with the steepest slope (maximum current sum). This is done by feeding the output of the zero-crossing circuits to a global WTA network with lateral excitation.[25] Lateral excitation accounts for the fact that the FOE position shifts smoothly in space: it facilitates the selection units close to the previously chosen winner and inhibits units farther away. In this way, once a strong zero-crossing is selected, the system will tend to track it as it moves in space. If high-contrast well-controlled stimuli are used, the WTA network in the last processing stage of the architecture is able to select and track in time the correct FOE location. For testing purposes, input stimuli were generated using a rotating drum with edges moving in a way to simulate *ego*-motion in a stationary environment. The stimuli were imaged onto the chip using a lens with an 8 mm focal length. Fig. 15 shows the output of the system, as the input stimulus is shifted from left to right.

4. Time to contact

Another quantity that can be extracted from an optical flow field is the *time to contact* with an object in the visual field. The time to contact is defined as the time it takes the observer to collide with the object, assuming that the relative velocity between observer and object remains constant. The time to contact is thus a very useful parameter for navigation systems, especially

in the case of *ego*-motion in a rigid environment. For translational motion with a relative speed v of the observer with respect to the object, the time to contact τ is given by

$$\tau = \frac{d}{v}, \tag{14}$$

where d denotes the distance between object and observer measured along the direction of relative motion. For translational motion towards a planar surface perpendicular to the optical axis of the imaging system, the velocity field $\mathbf{V}$ in the image is linear and its divergence $\nabla \cdot \mathbf{V}$ is thus constant across the surface. Using the 2-D version of Gauss's divergence theorem the time to contact can then be estimated robustly from the line integral of the normal velocity component along a closed contour.[26] If the contour is a circle C of radius r, we obtain

$$\tau = \frac{2\pi r^2}{\int_C \mathbf{V} \cdot \mathbf{n} ds}, \tag{15}$$

where $\mathbf{n}$ denotes the unit normal vector along the contour. Note that the focus of expansion does not have to lie within the contour and that its position, the relative speed v, and the distance d between object and observer do not have to be known. Furthermore, since eq. (15) depends on integrative rather than differential properties of the velocity field, the estimation of the time to contact is numerically stable, even in the presence of random noise and offsets, that are characteristic of subthreshold analog circuits.

Using the above result, the time to contact can be estimated with a circuit consisting of an array of 1-D velocity-sensing elements arranged on a circle, such that each element measures radial velocity. The line integral is then approximated by the properly normalized sum of all sensor outputs, so that the time to contact amounts to

$$\tau = \frac{N \cdot r}{\sum_{k=1}^{N} V_k}, \tag{16}$$

where N denotes the number of elements on the circle and V_k the radial velocity components at the locations of the elements.

We implemented such a circuit on a VLSI chip with 12 radially oriented FS velocity-sensing elements.[16] The layout of this chip is displayed in Fig. 16. The photodiodes of the velocity-sensing elements are arranged on two concentric circles with radii of 400 μm and 600 μm respectively. The output voltage of each element is used to control a subthreshold transistor current. Since this voltage is logarithmically dependent on velocity, the current is proportional to velocity and the sum of the velocity components can be calculated by aggregating the currents from all elements on two lines, one for outward motion and one for inward motion, and taking the difference of the total currents. The resulting bidirectional output current is then an inverse function of the signed time to contact.

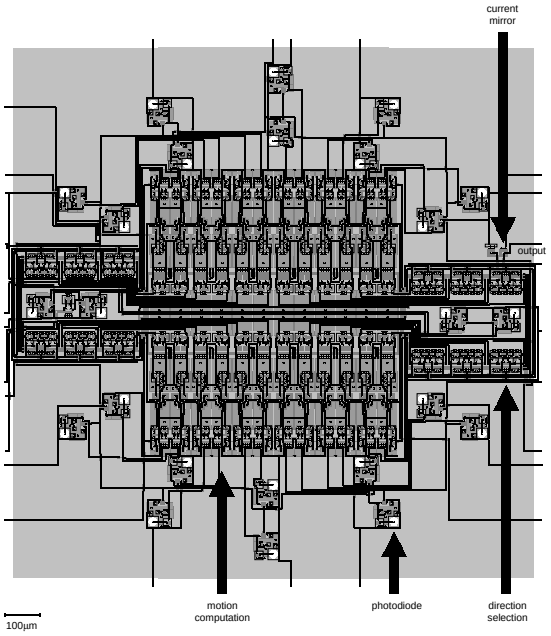


Fig. 16. Layout of an analog VLSI chip for the determination of the time to contact. The 12 pairs of photoreceptors arranged on two concentric circles with radii of 400 μm and 600 μm respectively are coupled to 12 FS velocity-sensing elements that estimate the radial components of the optical flow field. The pulse-shaping circuits and motion cells are located in the central part of the chip and the direction-selection circuits are located on the sides. The output voltages of the FS sensors are converted into currents, that are summed for each direction of motion (outward and inward) and subtracted (using a current mirror) to yield the final output current that is inversely proportional to the time to contact. The size of the layout is 1.6 mm $\times$ 1.6 mm.

The circuit yields reasonably accurate estimates of the time to contact for an approaching or receding pattern of high-contrast concentric rings centered on the focus of expansion and for a spiral stimulus on a rotating disk that simulates approaching or receding motion. Fig. 17 shows the averaged output current in response to the spiral stimulus as a function of simulated time to contact with a theoretical fit using eq. (16). The expected inverse relationship of output current and time to contact is qualitatively observed and the sign (expansion or contraction) is robustly encoded. However, the deviation of the output current from its average can be substantial. Since the output voltage of each velocity-sensing element gradually decays due to leak currents and since the spiral stimulus causes a serial update of the velocity values along the array, a step change in the output current

is observed upon each update, followed by a slow decay. The effect is aggravated, if the individual elements measure significantly differing velocities. This is generally the case, because the focus of expansion is often not centered on the sensor and because of inaccuracies in the velocity measurements due to circuit offsets, noise, and the *aperture problem*,[27] *i.e.* the fact that the direction of motion of a linear image feature cannot be determined from local measurements. Nonetheless, due to the integrative character of the algorithm, more robust results may be expected from stimuli with higher edge densities and arrays with a larger number of sensing elements, because statistical errors would be reduced and the high edge-detection rate would prevent the signal from decaying significantly or changing abruptly. In this way, good estimates for the time to contact in more general scenes should be obtained. To further improve accuracy, the chip could be mounted on a rotational stage that always centers it on the focus of expansion, using the information obtained from a system that computes the focus of expansion, such as the one described in the previous section.

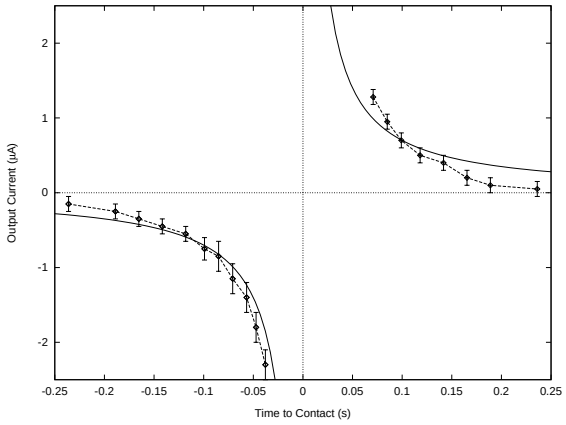


Fig. 17. Output current of the time-to-contact sensor versus simulated time to contact under AC-incandescent room illumination. The theoretical fit according to eq. (16) predicts an inverse relationship.

5. Motion discontinuities

Segmentation of images into different objects and a background is an important step in most image processing systems. In dynamic environments, objects can be segregated by their different apparent velocities with respect to the observer. These are either induced by motion parallax due to motion of the observer or by independent motion of the objects. For reasonably fast motions, segmentation based on *motion discontinuities* is less error-prone in complex environments than segmentation based on extracted edges. It therefore lends itself well to implementation in navigation systems mounted on rapidly moving platforms. Motion discontinuities can be extracted from an array of velocity-sensing elements by comparing the velocities measured by neighboring elements.

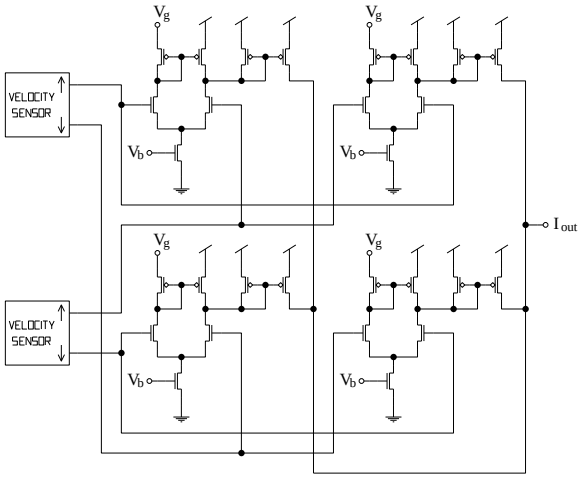


Fig. 18. Schematic diagram of the motion-discontinuity-detection circuitry for a pair of adjacent velocity-sensing elements. The absolute value of the difference of the output voltages is compared against a threshold (set by V_g) for both directions of motion. If the difference exceeds the set threshold for one or both directions, a current signal is output.

We implemented a circuit that finds motion discontinuities in a 1-D image and outputs current signals at their locations.[17] It uses a linear array of FS velocity sensors as a front end. The voltage outputs of pairs of adjacent elements are compared for both directions of motion separately. The schematic diagram of the circuit that performs this comparison is shown in Fig. 18. If the absolute value of the voltage difference for either direction of a pair exceeds a set threshold, a current is activated at the location of the pair, signaling a discontinuity. The value of this current is a monotonic function of the speed difference, that saturates at large differences. If the voltage difference remains below the threshold, the current remains shut off. The threshold is set with a bias voltage to exceed the fixed-pattern and temporal noise of the velocity-sensing array for uniform image motion. For testing purposes, the output currents at the different locations along the array are scanned off the chip using an on-chip scanner, passed through a linear current-to-voltage converter, and displayed on an oscilloscope. The array contains 24 FS sensors with a

pitch of 60 μm, giving 23 discontinuity measurements. The total size of the circuitry is 1.5 mm $\times$ 1.1 mm, as implemented with 2 μm technology. Fig. 19 shows a scope trace of the circuit response to a black bar translating in front of a background of uniformly moving black and white stripes. The two current peaks mark the locations of the edges of the bar. The velocity difference was large enough to saturate the output currents at the edge locations. The data of Fig. 19 was taken with the bar moving in the same direction as the background, but the same type of signal was measured for opposite directions of motion of bar and background. For uniform image motion no output was observed.

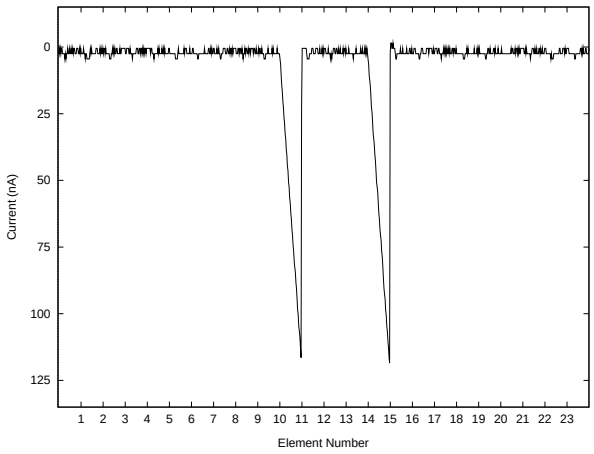


Fig. 19. Response of the motion-discontinuity chip to a black bar stimulus traveling across a striped background, that moves in the same direction at a different velocity. The velocities on chip were 5 mm/s for the bar and 30 mm/s for the background. The current peaks show the locations of the bar's edges on the imaging array. The output currents were scanned off the chip at a rate of 250 Hz. They were measured with a linear I-V converter, whose limited slew rate is responsible for the slow onset of the peaks.

If we feed a copy of the output signals of the velocity-sensing array into a resistive network to smooth out random noise and circuit offsets, we can use the motion-discontinuity signals to diminish or eliminate the amount of smoothing across object boundaries by increasing the local resistance or by opening switches at their locations.[28,29] With such an architecture it would be possible to implement smoothing and segmentation simultaneously. Unlike a resistive fuse network,[17,30] where discontinuity locations tend to persist in the network even after the input has changed, so that for a given stimulus multiple stable states may exist depending on the history of the stimulus, such a feedforward network would always yield a well-defined, unique solution.

6. CONCLUSIONS

We described three different types of analog VLSI velocity sensors that are compact enough to be implemented in focal plane arrays. They operate over velocity ranges of several orders of magnitude, while being reasonably invariant to illumination and contrast. We further presented single-chip systems based on one of these velocity sensors that compute the focus of expansion, time to contact, and motion discontinuities respectively, making use of integrative features of the velocity field which can be estimated robustly with circuits of limited precision. Such systems may prove useful for applications in vehicle navigation and robot guidance.

7. ACKNOWLEDGEMENTS

The FS and FT velocity sensors were developed in collaboration with R. Sarpeshkar. This work was supported by Daimler-Benz as well as by grants from the Office of Naval Research, the Center for Neuromorphic Systems Engineering as a part of the National Science Foundation Engineering Research Center Program, and by the Office of Strategic Technology of the California Trade and Commerce Agency. Fabrication of the integrated circuits was provided through the MOSIS prototyping service.

8. REFERENCES

[1] A. Verri and T. Poggio, "Motion field and optical flow: qualitative properties," *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 11, pp. 490-498, May 1989.

[2] T. Poggio, V. Torre and C. Koch, "Computational vision and regularization theory," *Nature*, Vol. 317, pp. 314-319, 1985.

[3] J. Tanner and C. Mead, "An integrated analog optical motion sensor," *VLSI Signal Processing*, Vol. II, S.-Y. Kung, R. E. Owen, J. G. Nash (eds.), pp. 59-76, IEEE Press, New York, NY, 1986.

[4] A. G. Andreou and K. Strohbehn, "Analog VLSI implementation of the Hassenstein-Reichardt-Poggio models for vision computation," *Proc. 1990 IEEE Int. Conf. Systems, Man, and Cybernetics*, pp. 708-710, Los Angeles, CA, Nov. 1990.

[5] A. G. Andreou, K. Strohbehn and R. E. Jenkins, "Silicon retina for motion computation," *Proc. 1991 IEEE Int. Symp. Circuits and Systems*, pp. 1373-1376, IEEE, Singapore, 1991.

[6] T. Horiuchi, J. Lazzaro, A. Moore and C. Koch, "A delay-line based motion detection chip," *Adv. Neural Inf. Proc. Systems*, Vol. 3, R. P. Lippmann, J. E. Moody, D. S. Touretzky (eds.), pp. 406-412, Morgan Kaufmann, San Mateo, CA, 1991.

[7] R. G. Benson and T. Delbrück, "Direction selective silicon retina that uses null inhibition," *Adv. Neural Inf. Proc. Systems*, Vol. 4, J. E. Moody, S. J. Hanson, R. P. Lippmann (eds.), pp. 756-763, Morgan Kaufmann, San Mateo, CA, 1992.

[8] R. Etienne-Cummings, S. Fernando, J. Van der Spiegel and P. Mueller, "Real time 2-D analog motion detector VLSI circuit," *Proc. Int. Joint Conf. Neural Networks*, Vol. IV, pp. 426-431, IEEE, Piscataway, NJ, 1992.

[9] T. Delbrück, "Silicon retina with correlation-based, velocity-tuned pixels," *IEEE Trans. Neural Networks.*, Vol. 4, pp. 529-541, May 1993.

[10] R. Etienne-Cummings, S. Fernando, N. Takahashi, V. Shtonov, J. Van der Spiegel and P. Mueller, "A new temporal domain optical flow measurement technique for focal plane VLSI implementation," *Proc. Comp. Arch. Mach. Perception*, M. Bayoumi, L. Davis, K. Valavanis (eds.), pp. 241-250, IEEE, Piscataway, NJ, 1993.

[11] R. C. Meitzler, K. Strohbehn and A. G. Andreou, "A silicon retina for 2-D position and motion computation," *Proc. 1995 IEEE Int. Symp. Circuits and Systems*, pp. 2096-2099, IEEE, Piscataway, NJ, 1995.

[12] S.-C. Liu, "Silicon model of motion adaptation in the fly visual system," *Proc. Third Joint Caltech/UCSD Symp. Neural Comp.*, Pasadena, CA, June 1996.

[13] J. Kramer, R. Sarpeshkar and C. Koch, "An analog VLSI velocity sensor," *Proc. 1995 IEEE Int. Symp. Circuits and Systems*, pp. 413-416, IEEE, Piscataway, NJ, 1995.

[14] J. Kramer, R. Sarpeshkar and C. Koch, "Pulse-based analog VLSI velocity sensors," accepted for publication in *IEEE Trans. Circuits and Systems II*, 1995.

[15] J. Kramer, "Compact integrated motion sensor with three-pixel interaction," *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 18, pp. 455-460, July 1996.

[16] G. Indiveri, J. Kramer and C. Koch, "Parallel analog VLSI architectures for computation of heading direction and time-to-contact," *Adv. Neural Inf. Proc. Systems*, Vol. 8, D. S. Touretzky, M. C. Mozer, M. E Hasselmo (eds.), pp. 720-726, MIT, Cambridge, MA, 1996.

[17] J. Kramer, R. Sarpeshkar and C. Koch, "Analog VLSI motion discontinuity detectors for image segmentation," *Proc. 1996 IEEE Int. Symp. Circuits and Systems*, Vol. II, pp. 620-623, IEEE, Piscataway, NJ, 1996.

[18] T. Delbrück, *Investigations of analog VLSI visual transduction and motion processing.* Ph.D. dissertation, Department of Computation and Neural Systems, California Institute of Technology, Pasadena, CA, 1993.

[19] R. Sarpeshkar, J. Kramer, G. Indiveri and C. Koch, "Analog VLSI architectures for motion processing: from fundamental limits to system applications," *Proc. IEEE*, Vol. 84, pp. 969-987, July 1996.

[20] J. Gibson, *The ecological approach to visual perception*, Houghton Mifflin, Boston, MA, 1979.

[21] G. Indiveri, J. Kramer and C. Koch, "Analog VLSI architecture for computing heading direction," *IEEE Proc. Intell. Vehicles Symp.*, pp. 24-29, IEEE, Piscataway, NJ, 1995.

[22] J. Lazzaro, S. Ryckebusch, M. A. Mahowald and C. A. Mead, "Winner-take-all networks of $O(n)$ complexity," *Adv. Neural Inf. Proc. Systems*, Vol. 1, D. S. Touretzky (ed.), pp. 703-711, Morgan Kaufmann, San Mateo, CA, 1989.

[23] K. A. Boahen and A. G. Andreou, "A contrast sensitive silicon retina with reciprocal synapses," *Adv. Neural Inf. Proc. Systems*, Vol. 4, J. E. Moody, S. J. Hanson, R. P. Lippmann (eds.), pp. 764-772, Morgan Kaufmann, San Mateo, CA, 1992.

[24] T. Delbrück, "'Bump' circuits for computing similarity and dissimilarity of analog voltages," *Proc. Int. Joint Conf. Neural Networks*, Vol. I, pp. 475-479, Seattle, WA, June 1991.

[25] G. Indiveri, "Winner-take-all networks with lateral excitation," accepted for publication in *J. Analog Integrated Circuits and Signal Processing*, 1996.

[26] T. Poggio, A. Verri and V. Torre, "Green theorems and qualitative properties of the optical flow," *MIT Internal Lab. Memo 1289*, 1991.

[27] A. Verri, M. Straforini and V. Torre, "Computational aspects of motion perception in natural and artificial vision systems," *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 11, pp. 490-498, 1989.

[28] C. Koch, J. Marroquin and A. Yuille, "Analog 'neuronal' networks in early vision," *Proc. Nat. Acad. Sci. USA*, Vol. 83, pp. 4263-4267, 1986.

[29] J. Hutchinson, C. Koch, J. Luo and C. Mead, "Computing motion using analog and binary resistive networks," *IEEE Computers*, Vol. 21, pp. 52-63, 1988.

[30] J. G. Harris, C. Koch and J. Luo, "A two-dimensional analog VLSI circuit for detecting discontinuities in early vision," *Science*, Vol. 248, pp. 1209-1211, 1990.