Compact Integrated Motion Sensor With Three-Pixel Interaction

Jörg Kramer

Abstract—An integrated circuit with on-chip photoreceptors is described, that computes the bi-directional velocity of a visual stimulus moving along a given axis in the focal plane by measuring the time delay of its detection at two positions. Due to the compactness of the circuit, a dense array of such motion-sensing elements can be monolithically integrated to estimate the velocity field of an image and to extract higher-level image features through local or global interaction.

Index Terms—Motion estimation, velocity sensor, optical flow, analog VLSI, robot vision.

1 Introduction

A variety of image-processing tasks, such as segmentation and estimation of depth, can be considerably simplified in dynamic scenes if motion data is available. Furthermore, mobile systems rely on motion information for the computation of important parameters of ego-motion, such as time to contact and focus of expansion. Since they use the extracted information to guide their behavior, processing has to be done within a time frame during which the scene does not evolve significantly, i.e., in "real time." Real-time image processing typically requires a very high bandwidth. In order to avoid a bottleneck after the photoconversion step, the input data is best handled in parallel by the first processing stages. Distributed preprocessing also makes the system more tolerant against failure of individual components and thus more reliable. Extraction of the information important to the system usually results in significant data reduction, making serial processing more attractive at later stages.

Mobile systems often place severe restrictions on size, weight, power consumption, and shock resistance of their image-processing circuitry, thus favoring highly-integrated solutions. However, the degree of integration is limited by the large size and power dissipation of parallel architectures. Well-designed analog systems are much more economic in area and power consumption than traditional digital ones with comparable computing power [1]. Since the high computational precision attainable with digital circuits cannot be fully exploited in systems with analog sensory inputs providing inaccurate and noisy data, the 8-bit accuracy typically offered by analog VLSI circuits may be quite appropriate for processing. For various systems, such rough estimates are sufficient to induce the proper action, especially if they provide feedback through ego-motion. Errors can be reduced by using *a priori* information, e.g., about the piecewise rigidity of the environment, to obtain some redundancy.

Algorithms for the computation of visual motion are usually classified into different categories [2], that are chosen on the basis of biological evidence or computational implications. For the purpose of this article, we shall distinguish between *gradient* and *correspondence* algorithms. Gradient schemes extract the velocity of an image feature from approximations of temporal and spatial derivatives of the local brightness distribution. Since the calculation of derivatives is sensitive to circuit offsets, noise, and varying irradiance, gradient algorithms are difficult to implement robustly with analog circuitry. Correspondence methods estimate motion by comparing the position of a pattern at different times (spatial correspondence), or by comparing the timing of a pattern at different positions (temporal correspondence). While digital implementations of correspondence algorithms typically use the former approach [3], most analog implementations and well-understood biological systems [4] do not intrinsically work on a discretization of time and thus naturally use the latter approach. Correspondence methods can be subdivided into *correlation* methods and *token-based* methods. Correlation methods operate on any type of image structure and hence, like gradient methods, produce a dense map of velocity estimates. However, they usually exhibit better numerical stability than gradient methods, since correlation is based on multiplication, rather than on differentiation. Token-based methods only respond to a particular class of image features, by first making a decision about their presence at a given location in space and time. At the expense of producing only sparse velocity maps, token-based methods can be made to operate quite robustly.

Given a host of fundamental difficulties in reliably estimating a velocity field from purely visual data (optical flow) [5] and the various trade-offs to be considered when designing a circuit to perform this task, the choice of the algorithm has to be guided by the specifications of the system and the environment it is used in.

Motion algorithms generally operate on discrete approximations of space and/or time, and thus suffer from aliasing effects. These are more serious than most other effects degrading the performance of a motion sensor, because they prevent the output signal from being a smooth function of the image parameters and therefore lead to completely spurious data instead of merely giving an inaccurate estimate of the velocity. Most implementations measure irradiance at discrete locations with a regular spacing Δx. This gives rise to a criterion for the prevention of spatial aliasing that sets an upper limit to the spatial frequency $\bar{k}$ in the image, given by the sampling theorem as

$$\bar{k} < \frac{1}{2\Delta x}. \tag{1}$$

Some analog correspondence algorithms operate on continuous time and are therefore not fundamentally limited any further by temporal aliasing. Most of these are token-based methods that infer the velocity v of an image token from its time of travel Δt between two adjacent locations according to

$$\Delta t = \frac{\Delta x}{v}. \tag{2}$$

Since the temporal frequency ν at a fixed location in the image is given by

$$\nu = \bar{k} \cdot v, \tag{3}$$

any stimulus that is not spatially aliased automatically meets the condition

$$\nu < \frac{1}{2\Delta t} \tag{4}$$

in such a time-of-travel algorithm. Gradient algorithms, digital correspondence algorithms and analog correlation algorithms, however, generally work with independent time constants Δt and are thus subject to a temporal-aliasing criterion of the form (4) that sets additional limitations onto the input signal.

Conditions (1) and (4) may be satisfied by low-pass filtering the input signal in space and in time respectively. Spatial low-pass

• *The author is with the California Institute of Technology, Division of Biology 139-74, Pasadena, CA 91125.*
E-mail: kramer@klab.caltech.edu.

Manuscript received May 15, 1995; revised Nov. 1, 1995. Recommended for acceptance by A. Singh.
For information on obtaining reprints of this article, please send e-mail to: transactions@computer.org, and reference IEEECS Log Number P95173.

0162-8828/96$05.00 ©1996 IEEE

filtering can easily be performed by optical means, whereas temporal low-pass filtering requires additional circuitry. The aliasing problem can thus be very conveniently solved with time-of-travel algorithms only requiring spatial low-pass filtering.

2 Relation to Previous Work

In the following, we give a short overview of previous implementations of analog VLSI motion sensors that combine photoreceptors and processing circuitry on a single chip. A more detailed survey can be found in [6]. We focus here on sensors of comparable size and performance to the one described in this article and assess the merits of this new scheme in relation to them.

After an early implementation of a 2D velocity sensor based on a gradient method [7] showed poor performance, subsequent efforts concentrated on implementing temporal-correspondence algorithms, ranging from pure correlation schemes [8] to algorithms performing correlation-type motion computation on extracted image tokens [9], [10], [11] and to token-based time-of-travel methods [11], [12], [13], [14]. Most of these previous analog VLSI motion sensors either only responded robustly to stimuli of very high contrasts [7], [11] or had an output signal that strongly depended on contrast and/or illumination, as well as velocity [8], [10], [12]. Another problem with some implementations [7], [9], [13] was the large size of the circuitry needed to extract a single velocity value, making them unsuitable for dense 2D focal-plane arrays.

Two token-based time-of-travel correspondence algorithms were implemented with compact circuits, that unambiguously encoded 1D velocity over considerable velocity, contrast, and illumination ranges. They both used the same feature extractor as an input stage, namely a "temporal edge detector" responding to dark-bright irradiance changes [14]. In the facilitate-and-trigger (FT) circuit [11], an edge signal generated a voltage pulse of fixed amplitude and length at each edge-detector location. The pulses from two adjacent locations were fed into two motion circuits, one for each direction. For motion in the preferred direction such a motion circuit output a pulse with a length that equalled the overlap time of the input pulses, while for motion in the null direction no output pulse was generated. In the facilitate-and-sample (FS) circuit [14], an edge signal gave rise to a sharp voltage spike and a logarithmically-decaying voltage signal at each detector location. The voltage spike from one location was used to sample the analog voltage of the slowly-decaying signal of an adjacent location, which was a measure of the relative time delay of the triggering of the two signals. Both circuits showed good performance for sharp edges of medium and high contrasts, whereby the output signal was independent of contrast and of the global illumination level over a considerable range [15]. For lower contrasts and more gradual edges the response started to decrease, thus underestimating the speed.

As published, all reviewed motion sensors were subject to a temporal-aliasing criterion in addition to the spatial-aliasing criterion, since besides fixed space constants they also used built-in time parameters for their computations, whose finite values were essential for their proper operation, thereby further limiting the working range in the temporal domain.

In this article we present a compact motion circuit based on a time-of-travel correspondence algorithm, that encodes velocity in the length of binary output pulses. The pulse length is inversely related to velocity, as for some other time-of-travel methods [11], [12], [13]. The use of active inhibition with a time constant adapting to the speed of the stimulus through a three-pixel-interaction scheme prevents temporal aliasing over a large velocity range. Throughout this article the proposed motion algorithm will be referred to as the facilitate-trigger-and-inhibit (FTI) algorithm. In the present implementation, the FTI motion cell receives its inputs from temporal edge detectors such as those used in conjunction with the FS and FT cells. With this input stage, the response characteristics of the FTI and FT sensors are similarly invariant against global illumination level and edge shape, whereas the output signal of the FS sensor shows more variance toward high velocities and low illumination and contrast levels. On the other hand, the FS cell provides good sensitivity over a large velocity range due to the logarithmic dependence of the output signal on the velocity, while for the FTI and FT schemes the inverse characteristics result in a quickly-decreasing sensitivity toward larger velocities. The FTI and FS sensors operate over a large velocity range, whereas the FT sensor does not respond to small velocities due to the finite length of the input pulses to the motion circuits. In contrast to the FT and FS cells, the FTI cell does not need any bias voltages in addition to those used by the edge detector.

Including the edge-detection stage, the total size of an FTI motion-sensing element in a 1D array, as implemented with 2 μm technology, is only about half that of such an FT or FS motion-sensing element. Without the edge detector, which may be implemented differently and may also be shared with circuits performing other types of computations, an FTI motion cell covers only about 20% of the chip area used by either an FT or an FS motion cell. Suppression of temporal aliasing in the FT and FS cells requires additional circuitry to compare the two directional outputs [15], making the size ratio even more favorable for the FTI cell.

Some of the implemented analog VLSI motion sensors aggregate data from sets of motion-processing elements or even from an entire array for a single velocity output [7], [8], [9], [10]. While on an algorithmic level, aggregation is only necessary for a few schemes [7], [10], it can be generally used to increase accuracy and reliability. For motion sensors with *value-coded* analog outputs, signal aggregation is straightforward to implement with current-summing techniques. Time-of-travel sensors, such as the FTI sensor, provide *time-coded* output signals, that have to be first temporally integrated before they can be aggregated with such methods. Alternatively, such signals can be directly averaged in the temporal domain using adaptation techniques. Such an approach is currently being evaluated for the FTI sensor.

3 Temporal Edge Detector

The FTI motion cell can be combined with any input stage that responds with a short current or voltage spike to a certain image feature. Detectors for low-level features, such as edges, provide denser feature maps and are simpler to build with analog integrated circuits than those for higher-level features. All reviewed implementations of token-based motion algorithms used edges as image tokens. Temporal edge detectors, responding to irradiance transients at a single position, are generally more compact and less prone to circuit offsets than spatial edge detectors, using irradiance data from different locations [16]. While being more sensitive to flicker noise induced by a.c.-driven light sources and other rapid global illumination transients, they can be operated at lower contrast-detection thresholds under slowly-varying lighting conditions. The FTI motion cell has been interfaced with the temporal edge detector described in [14], which was demonstrated to respond down to lower contrasts than the edge detectors used in most other reviewed circuits, while having a reasonably small size. It is schematically shown in Fig. 1.

The photoreceptor circuit (D_1, $Q_1 - Q_4$, $C_1 - C_2$) [17] consists of a sensor (D_1, Q_1) that converts the incoming irradiance into a voltage, a high-gain feedback circuit ($Q_2 - Q_3$), and a capacitive gain stage ($C_1 - C_2$) with slow adaptation (Q_4) of the output voltage V_{ph} to the steady-state response V_q, which increases logarithmically with irradiance E. Thus, the voltage transient dV_q caused by a

small irradiance transient dE, is only a function of the relative irradiance change dE/E and not of overall image brightness. The transient gain of the change dV_{ph} in the photoreceptor output voltage V_{ph} with respect to dV_q is given by $(C_1 + C_2)/C_2$.

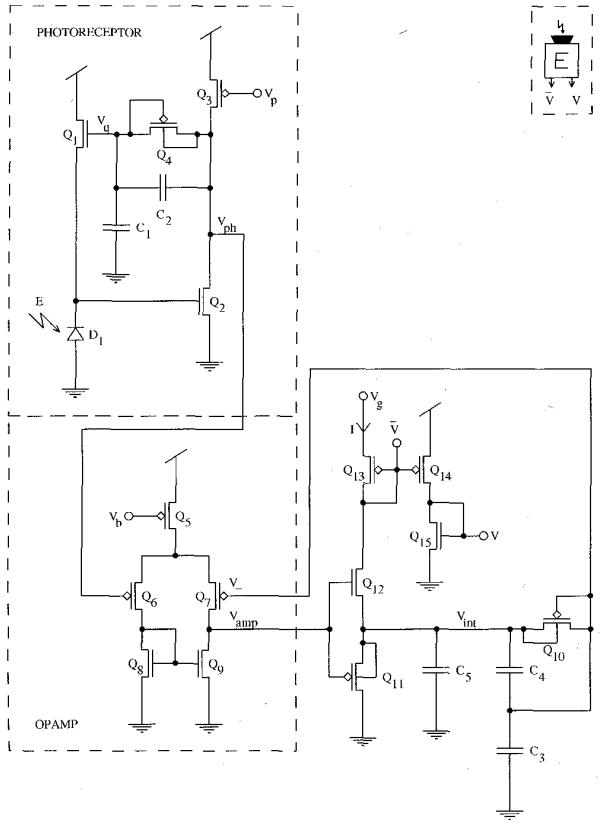

Fig. 1. Circuit diagram of a temporal edge detector [14]. A photoreceptor circuit (D_1, $Q_1 - Q_4$, $C_1 - C_2$) [17] transduces a transient dE in the incoming irradiance E into a voltage transient dV_{ph} that is proportional to the relative irradiance change dE/E. Additional circuitry ($Q_5 - Q_{15}$, $C_3 - C_5$) amplifies, differentiates, and rectifies this signal to provide output voltage spikes V and $\overline{V}$ in response to a rapid positive transient corresponding to a dark-bright or ON edge.

The voltage V_{ph} is fed into a circuit that converts a rapid positive transient dV_{ph} corresponding to a dark-bright or ON edge into a current spike I that is sensed by a voltage V and its complement $\overline{V}$. The circuit consists of a basic operational amplifier ($Q_5 - Q_9$), a capacitive gain stage ($C_3 - C_4$) with slow adaptation (Q_{10}), a differentiator (C_5), a rectifier ($Q_{11} - Q_{12}$), a current sensor (Q_{13}), and an inverter ($Q_{14} - Q_{15}$). During large irradiance transients the current I is given by

$$I = C_{tot} \cdot \frac{C_3 + C_4}{C_4} \cdot \frac{C_1 + C_2}{C_2} \cdot \frac{kT}{q\kappa} \cdot \frac{dE}{Edt} \cdot \Theta\left(\frac{dE}{dt}\right), \quad (5)$$

where

$$C_{tot} = C_5 + \frac{C_3 C_4}{C_3 + C_4} \quad (6)$$

and where kT/q denotes the thermal voltage, κ the back-gate coefficient, and Θ the Heaviside function. For further details of the circuit operation we refer to [14].

4 Motion Cell

The architecture of the FTI motion sensor is represented in Fig. 2. In a one-dimensional array of motion-sensing elements, each element contains an edge detector (E) and two motion circuits (M), each responding to one direction of motion. In Fig. 2 only the motion circuits of element i are shown. Unlike the previously-implemented motion algorithms reviewed in Section 2, most of which base their computations on two-pixel interactions, these motion circuits receive inputs from three neighboring edge detectors, at positions $i - 1$, i, and $i + 1$. The inputs to the motion circuits are the outputs V and $\overline{V}$ of the edge detectors (Fig. 1).

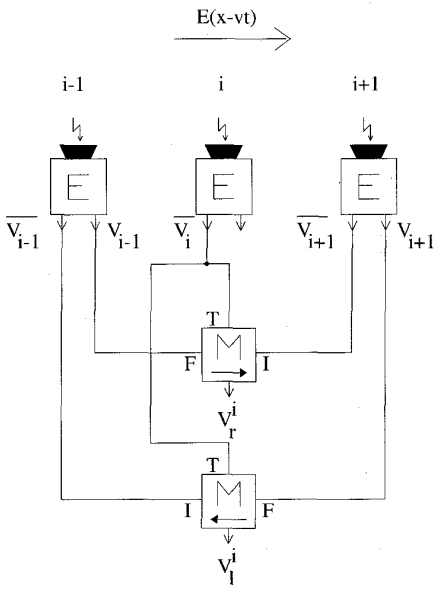

Fig. 2. Architecture of the facilitate-trigger-and-inhibit (FTI) motion sensor. Voltage spikes from edge detectors (E) at three adjacent pixel locations $i - 1$, i, $i + 1$ are used as facilitation (F), trigger (T), and inhibition (I) signals for output pulses V_r^i and V_l^i of direction-selective motion circuits (M). In a 1D array, each motion-sensing element consists of an edge detector and two motion circuits, one for each direction..

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 voltage spike V_{i-1}. Upon detection of the edge at position i, a voltage spike $\overline{V}_i$ is produced, that initiates 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 $\overline{V}_{i+1}$. The output-pulse length Δt is inversely proportional to the velocity v of the edge stimulus according to (2). If the edge moves from right to left, the voltage spike $\overline{V}_{i+1}$ is generated first and inhibits the triggering of an output pulse V_r^i by the succeeding voltage spike $\overline{V}_i$. The inhibition is terminated by the facilitation spike V_{i-1} when the edge is detected at position $i - 1$. The circuit for leftward motion has its facilitatory input connected to V_{i+1} and its inhibitory input to $\overline{V}_{i-1}$. It therefore only generates an output pulse V_l^i for edges moving from right to left.

This circuit does not use any internal time constants for the measurement of velocity. Unlike for some models for biological

systems [18] and other circuit implementations [11], [12], the facilitation and inhibition time intervals are neither built into the circuit nor set by external biases nor adjusted with feedback circuits. They 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. Facilitation in the preferred direction and inhibition in the null direction last twice as long as the output pulse. As pointed out in Section 1, the temporal-aliasing criterion (4) is thus equivalent to the spatial-aliasing criterion (1). The performance of the circuit in the temporal domain is therefore only limited by the time constants of the circuit components and not on an algorithmic level.

A schematic diagram of the two directional motion circuits is drawn in Fig. 3. The circuit computing rightward motion comprises transistors $Q_{16} - Q_{19}$ and capacitors $C_6 - C_7$. An output voltage pulse V_r^i is facilitated or inhibited depending on whether V_{inhr}^i is low or high. For motion in the preferred direction, a positive voltage spike V_{i-1} discharges C_7 through Q_{18}, turning off Q_{17} and therefore the inhibition. A negative voltage spike $\overline{V}_i$ then charges C_6 through Q_{16}, initiating an output pulse V_r^i. A negative spike $\overline{V}_{i+1}$ recharging C_7 through Q_{19} brings V_{inhr}^i to a high potential again, so that C_6 is discharged and the output pulse V_r^i is terminated. For motion in the null direction, $\overline{V}_{i+1}$ first turns on the inhibition, preventing an output voltage pulse to be triggered in response to a succeeding spike $\overline{V}_i$. The circuit for leftward motion ($Q_{20} - Q_{23}$, $C_8 - C_9$) is identical to the one for rightward motion, using V_{i+1} as facilitation spike and $\overline{V}_{i-1}$ as inhibition spike and generating a voltage pulse at V_l^i for motion in its preferred direction.

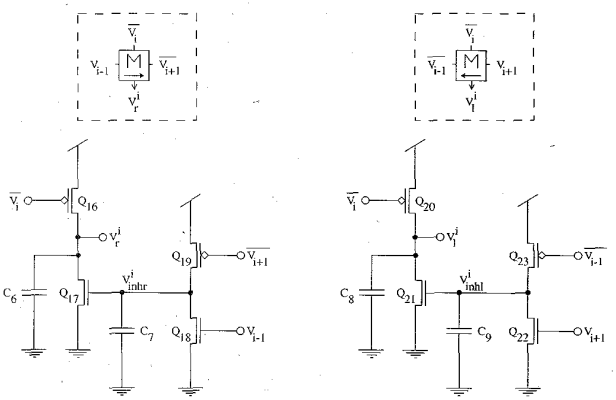

Fig. 3. Schematic diagrams of a pair of motion circuits responding to opposite directions of motion.

5 Experimental Data

A linear array of eight motion-sensing elements was fabricated using a 2 μm n-well CMOS process provided by MOSIS [19]. Each element comprises 23 MOSFETs and nine capacitors with a total area of 27,500 μm^2 of which the two motion circuits together, consisting of eight MOSFETs and four capacitors, cover 8,400 μm^2. The capacitors are small in order to conserve space and to provide short rise and fall times for the output pulses. The chosen pixel pitch Δx of 85 μm resulted in an element width of 323 μm. The object space was imaged onto the chip by means of a lens with a focal length f = 13 mm and an f-number of 1.8. For quantitative measurements, sheets of paper with printed gray-scale patterns wrapped around a rotating drum provided the optical stimuli. The object distance was set to 380 mm. Voltage traces from different stages of a motion-sensing element in response to a 56%-contrast ON edge are shown in Fig. 4. The data was taken under standard incandescent a.c. room lighting, where the white paper provided an illuminance of 1.2 lux onto the image plane. The output voltage pulse V_l for the preferred direction is enabled by a low value V_{inhl} during an adaptive time window, while the output voltage pulse V_r for the null direction is disabled by a high value V_{inhr} during that time window.

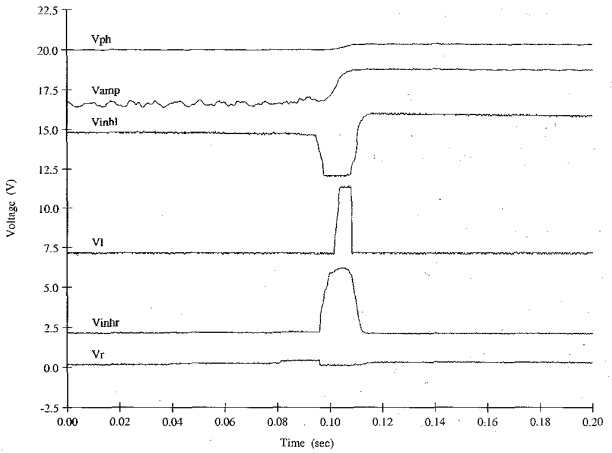

Fig. 4. Output voltage traces of different stages of a motion-sensing element in response to a moving 56%-contrast ON edge under incandescent a.c. room illumination. V_{ph} and V_{amp} are voltage traces of the edge detector at the outputs of the photoreceptor and of the operational amplifier respectively. V_{inhl} and V_{inhr} show the voltages at the inhibition nodes of the motion circuits for the preferred and null direction respectively. V_l and V_r are the output voltages for the preferred and null direction respectively. For better visibility, the voltage traces for V_{ph}, V_{amp}, V_{inhl}, V_l, and V_{inhr} are offset by 19 V, 16 V, 12 V, 7 V, and 2 V, respectively. The pixel spacing of 85 μm corresponds to an optical angle of 0.36° and a distance of 2.4 mm on the object. The pulse length of V_l is 6.2 msec, corresponding to an image speed of 13.7 mm/sec.

The output-pulse length for the preferred direction as a function of the velocity in the image for the same stimulus is plotted in Fig. 5. The response curves for three adjacent elements and a theoretical curve as obtained from (2) are shown. Each data point represents the average of eight measurements. The measured data agrees well with the theoretical prediction with a fixed-pattern noise low enough to make the fabrication of larger arrays seem practical. In the null direction, no output pulses were observed for the plotted speed range. For speeds higher than 60 mm/sec, however, the circuit responds in the null direction as well. This is not due to temporal aliasing (i.e., trigger and inhibition spikes being caused by different edges), but to the finite rise time of the inhibition signal, which only becomes completely effective after the edge has reached the trigger location and the output voltage has already started to increase. The rise time of the inhibition signal is determined by the shape and amplitude of the voltage pulse from the edge detector (and thus by the sharpness and the contrast of the edge) and by the size of the capacitor at the inhibition node. A circuit that provides shorter rise times, using three additional transistors in the edge detector, is currently being tested. At the low end, the velocity range is limited by the decay-time constants, due to leakage currents, of output and inhibition pulses. For the used stimulus, inhibition in the null direction and a significant output-pulse amplitude in the preferred direction can be maintained for approximately 2.5 sec, corresponding to a speed of 0.034 mm/sec.

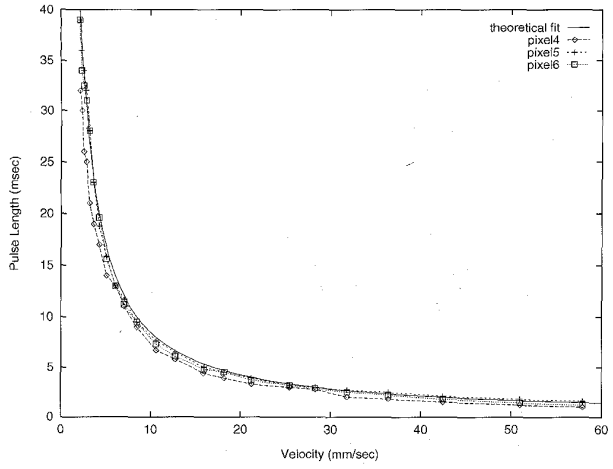

Fig. 5. Output-pulse length of three adjacent motion-sensing elements for the preferred direction of motion of a 56%-contrast ON edge versus image velocity for incandescent room illumination with theoretical fit according to (2). Each data point represents the average of eight successive measurements.

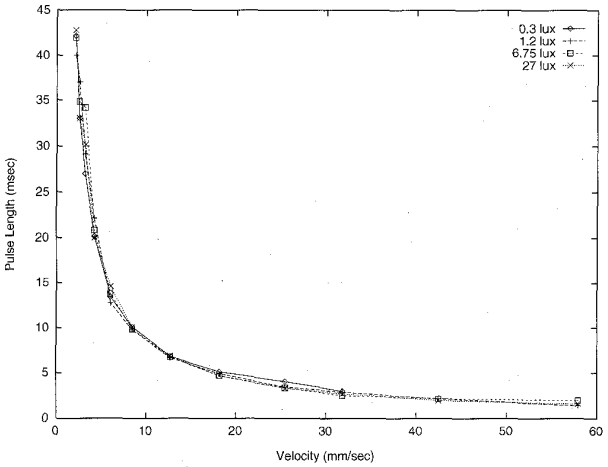

Fig. 6. Output-pulse length of a motion-sensing element for the preferred direction of motion of a 56%-contrast ON edge versus image velocity for different levels of incandescent illumination. Each data point represents the average of six successive measurements.

The response of a motion-sensing element to the same stimulus under different incandescent illumination levels is shown in Fig. 6. For each data point, six measurements were averaged. The response is very robust against changes in global light conditions over at least two orders of magnitude. The effect of changing edge contrast and sharpness is shown in Fig. 7. Contrast is defined here as the ratio of the difference and the sum of maximum and minimum irradiance. The response to sharp edge stimuli with different contrasts is compared with the response to a pattern whose gray value varies sinusoidally from dark to bright across a distance of 75 pixels. Six measurements taken under an illuminance of 1.2 lux were averaged per data point. For sharp edges with medium or high contrasts the response is insensitive to contrast variations. For very blurred edges the averaged output-pulse length still depends monotonically on the velocity, but the response curve shifts slightly toward longer pulses, leading to an underestimation of the speed. The same qualitative behavior was observed for sharp edges with low contrasts. The slow positive irradiance transient caused by the sinusoidal stimulus occasionally triggered multiple edge signals at some pixels, leading to spurious long pulses representing a low speed in the null direction. The edge detector was operated at low gain with V_b above threshold to prevent the brightness modulations of the incandescent illumination from generating spurious output spikes. Under d.c.-lighting conditions the gain can be increased, extending the contrast invariance of the output signal down to lower contrasts.

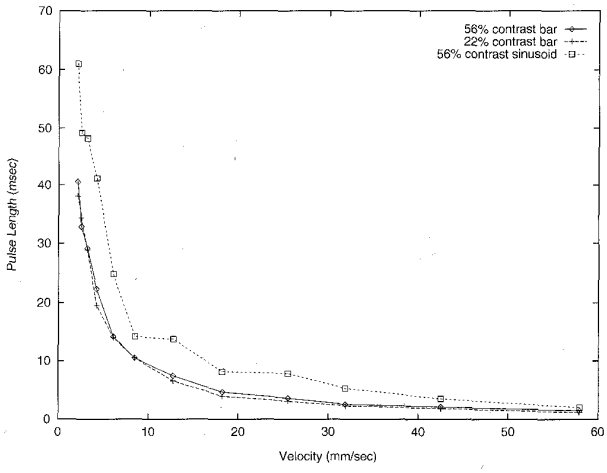

Fig. 7. Output-pulse length of a motion-sensing element for the preferred direction of motion of different ON edges for incandescent room illumination. Each data point represents the average of six successive measurements.

6 Conclusion

We presented an algorithm for the real-time measurement of bidirectional velocity of an optical stimulus in one dimension and its implementation with analog circuitry using standard CMOS technology. The circuit is more compact than that of most previously-reported optical velocity sensors. Its output is a monotonic function of velocity and is independent of the global illumination level across at least two orders of magnitude and of edge shape for sufficiently sharp and high-contrast edges. For more gradual and lower-contrast edges the response starts to shift toward lower speeds. Temporal aliasing is prevented by using an adaptive inhibition scheme. Dense arrays of such motion sensors can be monolithically integrated for the acquisition of entire velocity maps, which can be used to obtain higher-level image descriptions suitable for navigation of mobile systems in complex environments.

Acknowledgments

This work was supported by grants from the Swiss National Science Foundation, 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.

References

[1] C. Mead, "Neuromorphic electronic systems," *Proc. IEEE*, vol. 78, pp. 1,629-1,636, Oct. 1990.

[2] E.C. Hildreth and C. Koch, "The analysis of visual motion: From computational theory to neuronal mechanisms," *Ann. Rev. Neuroscience*, vol. 10, pp. 477-533, 1987.

[3] C. Koch, A. Moore, W. Bair, T. Horiuchi, B. Bishofberger, and J. Lazzaro, "Computing motion using analog VLSI chips: An ex-

perimental comparison among four approaches," *Proc. IEEE Workshop Visual Motion*, pp. 312-324, Princeton, N.J., Oct. 1991.
[4] B. Hassenstein and W. Reichardt, "Systemtheoretische Analyse der Zeit-, Reihenfolgen und Vorzeichenauswertung bei der Bewegungsperzeption des Rüsselkäfers *Chlorophanus*," *Z. Naturforsch.*, vol. 11b, pp. 513-524, 1956.
[5] A. Verri and T. Poggio, "Motion field and optical flow: Qualitative properties," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, no. 5, pp. 490-498, May 1989.
[6] R. Sarpeshkar, J. Kramer, G. Indiveri, and C. Koch, "Analog VLSI architectures for motion processing: From fundatmental limits to system applications," submitted to *Proc. IEEE*, Aug. 1995.
[7] J. Tanner and C. Mead, "An integrated analog optical motion sensor," *VLSI Signal Processing, II*, S.Y. Kung, ed., pp. 59-76. New York: IEEE Press, 1986.
[8] A.G. Andreou, K. Strohbehn, and R.E. Jenkins, "Silicon retina for motion computation," *Proc. 1991 IEEE Int'l Symp. Circuits and Systems*, pp. 1,373-1,376, Singapore, June 1991.
[9] T. Horiuchi, J. Lazzaro, A. Moore, and C. Koch, "A delay line based motion detection chip," *Advances in Neural Information Processing Systems 3*, R. Lippman, J. Moody, D. Touretzky, eds., pp. 406-412. San Mateo, Calif.: Morgan Kaufman, 1991.
[10] T. Delbrück, "Silicon retina with correlation-based, velocity-tuned pixels," *IEEE Trans. Neural Networks*, vol. 4, pp. 529-541, May 1993.
[11] R. Sarpeshkar, W. Bair, and C. Koch, "Visual motion computation in analog VLSI using pulses," *Advances in Neural Information Processing Systems 5*, D.S. Touretzky, ed., pp. 781-788. San Mateo, Calif.: Morgan Kaufman, 1993.
[12] R.G. Benson and T. Delbrück, "Direction selective silicon retina that uses null inhibition," *Advances in Neural Information Processing Systems 4*, D.S. Touretzky, ed., pp. 756-763. San Mateo, Calif.: Morgan Kaufman, 1991.
[13] 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. Computer Architecture for Machine Perception 1993*, M. Bayoumi, L. Davis, and K. Valavanis, eds., pp. 241-250, 1993.
[14] J. Kramer, R. Sarpeshkar, and C. Koch, "An analog VLSI velocity sensor," *Proc. 1995 IEEE Int'l Symp. Circuits and Systems*, pp. 413-416, Seattle, May 1995.
[15] J. Kramer, R. Sarpeshkar, and C. Koch, "Pulse-based analog VLSI velocity sensors," submitted to *IEEE Trans. Circuits and Systems II*, Apr. 1995.
[16] W. Bair and C. Koch, "An analog VLSI chip for finding edges from zero-crossings, *Advances in Neural Information Processing Systems 3*, R. Lippman, J. Moody, D. Touretzky, eds., pp. 399-405. San Mateo, Calif.: Morgan Kaufman, 1991.
[17] T. Delbrück, "Investigations of analog VLSI visual transduction and motion processing," PhD dissertation, Dept. of Computation and Neural Systems, California Inst. of Technology, Pasadena, Calif., 1993.
[18] H.B. Barlow and W.R. Levick, "The mechanism of directionally selective units in the rabbit's retina," *J. Physiology*, vol. 178, pp. 447-504, 1965.
[19] MOSIS (MOS Implementation System), Information Sciences Inst. of the Univ. of Southern California, Marina del Rey, Calif.

Global Word Shape Processing in Off-Line Recognition of Handwriting

Christophe Parisse

Abstract—Off-line recognition of handwriting may be achieved using simplified *profiles* of word shapes. These profiles consist of approximations of the word's upper and lower contour. Training and recognition are based on n-gram extraction and identification. The lexicons used extend to 16,000 words.

Index Terms—Handwriting, off-line, global recognition of word shape, contour, n-gram, large-size lexicon.

1 Introduction

FROM the time of the earliest work on the off-line machine recognition of handwriting, the problem arose of the choice between a global approach intended to process entire words [5] and an analytic one consisting of processing words on a letter-by-letter basis [14]. The letter-by-letter approach which has undergone the most development [1], [2], [3], [8], [13] may not be applicable in the case of unconstrained poor quality script where one is regularly confronted with important deformations or the absence of one or several letters. These local distortions rarely modify the overall word shape, and herein lies the interest of global recognition procedures. Yet, due to the complexity of global shapes, work on the global recognition of handwritten words remains rare [7], [9], [15].

The aim of this article is to propose solutions to this difficulty. The image of a word is very complex and its global shape does not reflect all the local accidents of the shapes of the letters constituting the word. Thus, a handwritten word formed of letters clumped and simplified to the extreme almost always contains zones which, detached from the context of the word, would no longer be identifiable. To read such words, the human reader is lead to find a sort of prototypical shape of the complete word. The notion of prototype suggests that exhaustive processing of all the information of the script is not required, especially local information which is often the most variable. In this sense a prototype may be viewed as a sort of simplification of the global image which may be directly compared to other simplified images. Even very rough characteristics of the word's image permit an approximate identification of the word [11]. However, such a global method does not allow for the differentiation of words that are visually very close and should be complemented later by local verifications to achieve precise recognition [13], [6].

This article is divided into three parts. The first part is devoted to techniques of approximation (simplification) of word shapes, designed to bring out the most stable elements. This stability (linked to the goal of the system) is the general look or appearance of the word. This notion of appearance may be found in all research using features like ascenders or descenders from the body of the word [4], as well as in research based on the notion of contour [17], [20]. It is on the basis of these two notions that it was possible to derive the concept of the upper and lower contour of a

• *The author is with INSERM-TLNP, Pavillon Claude Bernard, Hôpital de la Salpétrière, 47 Bd de l'Hôpital, 75651 Paris Cedex 13, France. E-mail: parisse@idf.ext.jussieu.fr.*

Manuscript received July 14, 1994; revised Aug. 14, 1995. Recommended for acceptance by M. Mohuiddin.

For information on obtaining reprints of this article, please send e-mail to: transactions@computer.org, and reference IEEECS Log Number P95156.

0162-8828/96$05.00 ©1996 IEEE