Computational and Implementation Strategies for Smart Visual Sensors in Automotive Applications

F. Solari, S.P. Sabatini, and G.M. Bisio
PSPC Research Group-DIBE University of Genoa
Via all'Opera Pia 11A, 16145 Genova, ITALY e_mail: fabio@dibe.unige.it

1 INTRODUCTION

In industrial and consumer worlds there are many alerting and guidance tasks that could be performed by peripheral vision sensors based on dynamic measurements over gross-feature space: (a) low-end products for every-day life visual requirements (e.g., such sensors could be used as rear-view-mirrors in a car safety-system, or as detectors for discriminating movement of infants/patients at home or at health-care centres); (b) vision aid for elderly and handicapped people; and (c) in general, *behavioral systems* able to provide a *reflex* as a shortcut to action (see Fig. 1). In all these various situations, a preattentive control mechanism must be active to handle unexpected events, possibly combined with habituation to reduce the impact of repetitive stimuli. In this scenario, a growing attention has been paid to analog perceptual solutions in which the system architecture integrates the image sensor with the preprocessing circuitry (cf. silicon retinas) [1]. This approach is particularly attractive for those early-vision applications in which flexibility of computation (otherwise, essential in a general-purpose vision system) is less important than the speed to go from signals to action. In such devices, information processing occurs at the physical level, i.e., sensorial signals are transduced into new physical variables suitable for the system and it is the evolution of the system that performs computation, carried out massively in parallel with high efficiency and speed.

The paper reports on the status of ongoing research aimed at the development of analog VLSI sensory modules with embedded dynamic perceptual processing. The computational principles of these dedicated VLSI modules are studied on the assumption that a most efficient implementation of these perceptual tasks can be based on *analog structured arrays reacting collectively to spatio-temporal input stimuli*, interpreted as the collective behavior of reaction and diffusion (RD) systems. Accordingly, early vision perceptual processes are interpreted as a "measuring" operation on the visual signal and local relational operators are introduced to extract structural properties of visual data (orientation, texture, motion, etc.) by analyzing a single point of the visual field in terms of *how* it is related to what is around it, in space and time. The paper is organized as follows. First, to characterize the essence of a discrete RD network, a basic cooperative block (the "Perceptual Engine") is introduced, and its practical realization, performance and limitations are described. Secondly, its application in the hardware implementation of an important class of perceptual problems, such as stereo disparity estimation, and motion analysis, is considered.

2 REACTION AND DIFFUSION SYSTEMS AS A PERCEPTUAL MEDIUM FOR EARLY VISION

The preattentive part of the visual system makes a crude analysis of the visual field through spatio-temporal measurements on the gross-feature space. The low-level cues so acquired (e.g., visual novelty, peripheral motion), together with expectation and intermodal (e.g., audition) stimulation, drive the attentional shifts towards new points-of-interest (i.e., fixation points) in a continuous process guiding

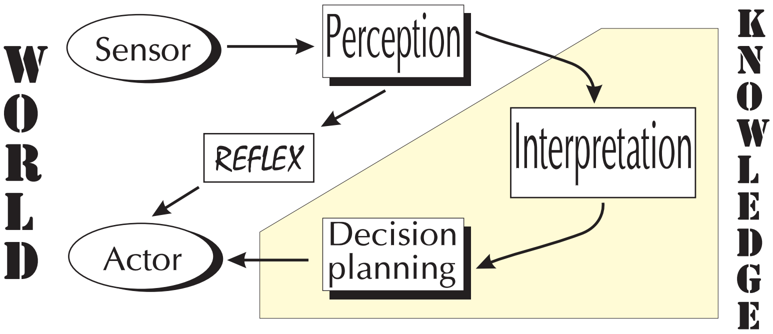


Figure 1: Schematic organization of vision guided behavior.

behavior [2]. The broad field of view provided by peripheral vision has both alerting and guidance functions. With respect to foveal vision, peripheral vision has enhanced dynamic performances [3] used for motion detection and for providing the motor control with the information necessary to foveate over changes in the visual field [4]. So, the appearance of a peripheral stimulus leads to a strong temporal cue often inducing a movement of gaze toward the stimulus.

Technically, we can observe that joint gradients in space and time are the most important features for preattentional shifts [2]. Accordingly, to emulate the functionality of peripheral vision one has to resort to those computational paradigms where spatial and temporal information interplay in generating perceptual output. In this respect, RD systems appear to be a most fitting alternative. In most general terms, RD systems are "an assembly of a large number of identical local systems which are coupled (i.e., diffusion coupled) to each other [5]"; this coupling is representative of the cooperative actions which take place among local systems. Formally, a RD process can be written as the solution of a system of coupled partial differential equations:

$$\frac{\partial \mathbf{e}}{\partial t} = \mathbf{A}(\mathbf{e}) + \mathbf{B}(\mathbf{s}) = \mathbf{D}\Delta\mathbf{e} + \mathbf{F}(\mathbf{e}) + \mathbf{B}(\mathbf{s}) \tag{1}$$

where $\mathbf{e}$ is the spatiotemporal excitation, $\mathbf{s}$ is the external stimulus, $\mathbf{D}$ is the diffusion matrix, Δ is the Laplacian operator in the spatial coordinates, $\mathbf{B}$ is the input connection vector, and the vector $\mathbf{F}$ is a catch-all term describing all reactions and interactions [6]. RD systems are natural candidates to model biological behavior since their diffusion component allows for a weighted averaging of the input signal over a region of a limited extent (e.g., the receptive field), thus realizing a reduction in the complexity of the signal, while the reaction components (possibly non linear) provide a flexible medium through which all sorts of excitatory and inhibitory behavior can be modeled.

2.1 Lattice models for early vision

The discrete forms of RD operators (i.e., lattice network) make it possible to map them on a layered network of interacting nodes. The network gains its system functionality through the cooperative action of ensembles of nodes topologically organized. Assuming linear reactions, the corresponding steady-state ($\partial e/\partial t = 0$) system can be described through the following input/output relation:

$$\sum_{k,l} a_{kl}\, e(m-k, n-l) = \sum_{k,l} b_{kl}\, s(m-k, n-l) \tag{2}$$

where $s(m,n)$ and $e(m,n)$ are the input and the output signals of the processing node at location (m,n), and the sets I and O determine the (rectangular) neighborhood domains of the feedforward excitation and cooperative interactions, respectively. Eq. 2 can be interpreted as a 2D iterative filter that detects specific characteristics present in the input pattern of excitation. The spatial domain on which these characteristics are detected is possibly larger than that of physical interconnections; this occurs both directly, by local interactions within the local neighborhood O, and indirectly, through the *propagation property* of iterative filters, which asserts that the output values of the filter can be recurrently affected by a larger neighbor region on the lattice network. As the system is linear and space-invariant, the output distribution (e) can be determined by the convolution between the input distribution (s) and the impulse response $h(\cdot,\cdot)$:

$$e(m,n)=\sum_k\sum_l h(m-k,n-l)s(k,l). \tag{3}$$

The shape and size of the kernel $h(\cdot,\cdot)$ depend on the sets of parameters a's and b's as can be explicited by introducing the Z transform of the impulse response.

In this paper, we restrict our analysis to filters for which the kernels are separable:

$$e(m,n)=\sum_k h_1(m-k)\sum_l h_2(n-l)s(k,l). \tag{4}$$

Under these conditions, the 2D problem can be recast as the cascade of two 1D problems.

2.2 The Perceptual Engine

An interesting case study is represented by an 1D diffusive layer with linear recurrent inhibitory reactions. Given the difference equation describing diffusive nearest couplings and next nearest neighbors inhibitory reactions among nodes:

$$a_{-2}e(n-2)+a_{-1}e(n-1)+a_0e(n)++a_1e(n+1)+a_2e(n+2)=b_0s(n) \tag{5}$$

one derives the Z-domain transfer function and obtains the spatial impulse response [7]:

$$h(n)=C\mathrm{e}^{-\lambda|n|}((1+\mathrm{e}^{-2\lambda})\sin(\omega_0)\cos(\omega_0 n)+(1-\mathrm{e}^{-2\lambda})\cos(\omega_0)\sin(\omega_0|n|)) \tag{6}$$

where $b_0=C[(1-\mathrm{e}^{-4\lambda})(1+\mathrm{e}^{-2\lambda})\sin(\omega_0)-4\mathrm{e}^{-2\lambda}\,(1-\mathrm{e}^{-2\lambda})\sin(\omega_0)\cos^2(\omega_0)](1-\mathrm{e}^{-4\lambda})$, $a_{-2}=a_2=\mathrm{e}^{-2\lambda}$, $a_{-1}=a_1=-2\mathrm{e}^{-\lambda}(1+\mathrm{e}^{-2\lambda})\cos\omega_0$, $a_0=1+4\mathrm{e}^{-2\lambda}\cos^2\omega_0+\mathrm{e}^{-4\lambda}$. In the discussion that follows we shall consider the $h(\cdot)$ function as a primary characteristic of a network rather than a derived quantity to be obtained by the solution of a difference equation. This function ($h(n)$) can be interpreted as the "Perceptual Engine" of our systems, providing computational primitives that can be composed to obtain more powerful image descriptors. Specifically, it can be proven that a local weighted sum over the Perceptual Engines leads to periodic functions as the result of an amplitude modulation of a sine plane wave with an exponential envelope:

$$g(n)=\alpha h(n-1)+\beta h(n)+\gamma h(n+1)=C\mathrm{e}^{(-\lambda|n|)}\cos(\omega_0 n+\varphi). \tag{7}$$

where $\alpha=C\mathrm{e}^{-\lambda}(\mathrm{e}^{-2\lambda}-1)\cos(\omega_0-\varphi)$ $\beta=C(1-\mathrm{e}^{-4\lambda})\cos(\varphi)$ $\gamma=C\mathrm{e}^{-\lambda}(\mathrm{e}^{-2\lambda}-1)\cos(\omega_0+\varphi)$. These functions closely resemble in space the Gabor functions which are extensively used in computational vision due to their optimal conjoint spatial/spatial-frequency representation property.

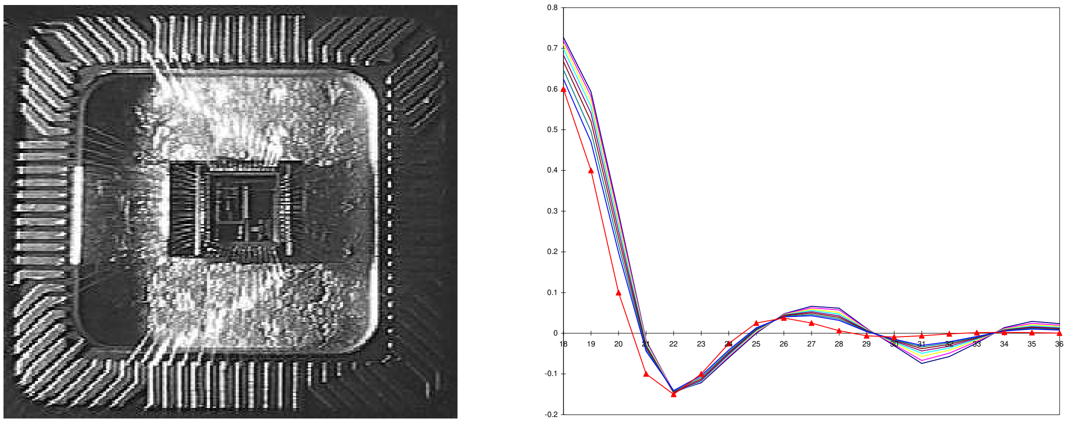
If one introduces a temporal dependence in the weighting coefficients α, β and γ of Eq. (7), it is possible to obtain a spatio-temporal linear filter. Specifically, we can rewrite Eq. (7) as:

$$g(n,t)=C'\mathrm{e}^{-t/\tau}\mathrm{e}^{(-\lambda|n|)}\cos(\omega_0 n+\omega_0 v_f t). \tag{8}$$

having posed $C(t)=C'\mathrm{e}^{-t/\tau}$ and $\varphi(t)=\omega_0 v_f t$, where the parameters τ and v_f characterize the temporal response of the second-order network generating $\alpha(t)$, $\beta(t)$, and $\gamma(t)$. In this way, we obtain a 1D spatio-temporal Gabor-like kernel with a linearly-variable phase interpretable by a dumped localized traveling wave with constant velocity v_f. If we stimulate the system with a moving spot of light, with speed v_p, its temporal response will be the envelope of the Gabor-like function only if $v_f=v_p$. Thus, such spatio-temporal linear filter will be able to detect the presence or the absence of an impulsive input moving along a specific direction with speed $v_p=v_f$.

2.3 Distributed computation on analog physical systems

Analog recurrent filter operations can be implemented by using linear resistive networks with positive and negative resistors [8] [7], or by current generators driven by the voltage at other nodes (VCCS) [9]. Since our main interest is to investigate how complex functionalities can be achieved through cooperative computations, we choose to relate the lattice networks' equations to a circuit architecture of one-way-interacting elements implemented as current controlled current sources (CCCS). Let us thus consider a regular grid of nodes, in which each node is fed by a current generator whose value is proportional to the incident light at that point (feedforward contribution). By using a current mode technique, in which all signals are encoded by currents, the excitation e of the network is mapped on a current, the interaction can be implemented by CCCSs that feed or sink currents according to the current values of neighboring nodes. In this way, the analog signal processing of our circuit architecture will be based on ratios of matched component values [10], thus eliminating the dependence on the performances of single devices. The sum of the current contributions from other nodes is performed by Kirchhoff current law at the node. A prototype 36-node VLSI circuit was fabricated on CMOS 0.8 μm, N-well, single poly, and double metal technology (see Fig. 2a). Figure 2b compares measured and expected impulse response for



(a) (b)

Figure 2: (a) Microphotograph of the 1D 36-node chip. (b) Responses of the chip measured for different values of the input current injected in the central node. The theoretical response is superimposed as the triangle dot curve. Only the right half nodes response is evidenced.

different values the injected current in the central node. It is worthwhile noting that even though the match between the expected response (the line with triangular dots) and the measures is not exact, the output corresponds in any case to a Perceptual Engine with slightly different parameters.

3 System architecture

The first stages of a vision system should be able to extract quickly and efficiently the *salient* information in the image, represented in natural scenes by lines, edges, textures, stereo cues, velocity fields, etc. Various computational vision models can make use of Gabor-like functions as localized spatial filters or as basic functions for image transformations [11]. Accordingly, the Perceptual Engine can be used as an elementary operator on its own, or as a building block to obtain more elaborate visual operators. The analog microsystem can be integrated within a board through different control and post-processing modules, to obtain a stand-alone system. The characteristics of the post-processing modules depend on the specific tasks (e.g. stereo depth estimation, motion analysis) to be performed. Fig. 3 shows the general architecture for a machine vision system based on the analog perceptual microsystem. In this

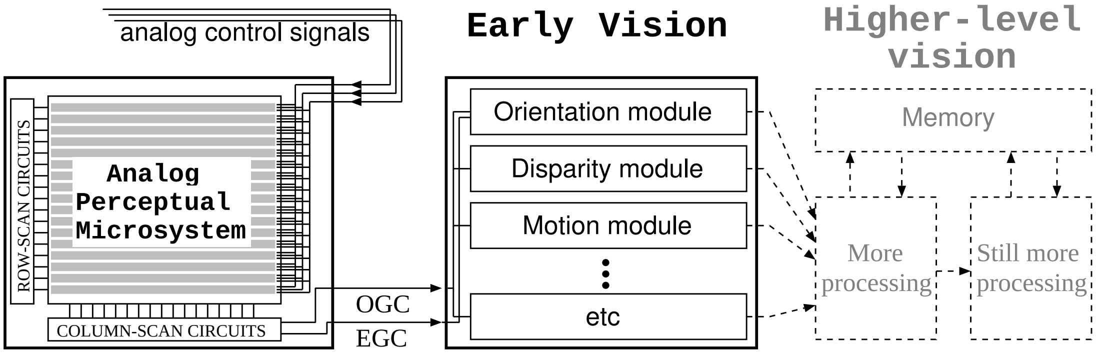


Figure 3: General architecture for a machine vision system based on the analog perceptual microsystem. A rank of 1D Perceptual Engines makes up the analog core of the architectural system: 1D node arrays are placed in rows and receive the excitation signals from sensor outputs. The column-scan circuits perform the local weighted combination of the output currents coming from the chip core to obtain the pair of quadrature Gabor filters (OGC and EGC). The Perceptual Engine kernel could be possibly controlled in frequency and spatial extension through external analog control signals. Subsequent processing are performed on the OGC/EGC signals, by further modules integrated on the same board in a hybrid analog/digital subsystem.

section, to prove the application capabilities of such microsystem as a front-end module in a machine vision system, we examine, as an example, its potential application in stereoscopic depth perception.

Stereopsis Stereoscopic depth estimation is mainly based on disparity measurement. Disparity is a local property that can be associated to a stereo pair (left and right) of images or image sequences. This problem has an elegant formulation in the frequency domain, in relation to the so-called phase-based matching methods [12]. Basically, disparity estimation is computed as the shift necessary to align the phase values of the band-pass filtered version of the binocular stereo signals. One can assume that the disparity between the stereo image pair occurs only along the direction of the epipolar lines [13]. In the simplest situation, in which optical axes are parallel, epipolar lines are parallel to the line joining the optical centers of the cameras. In such conditions, stereo phase-based matching can be cast as a 1D problem. Hence, we can use a rank of 1D Perceptual Engines acting separately on different rows of the input image. The local phase is computed following Sanger's algorithm [12], from the output of a pair of Gabor filters in quadrature (see Fig. 3).

The functional organization of the stereoscopic perception system is sketched in Fig. 4: the differential measure of the local phase, obtained by the left and right retinas, is validated by a confidence measure, thus deriving a reliable disparity map. In Fig. 5 the image reproduces a simple toy-car reconstruction of a traffic/vehicle scene. The disparity map, obtained through a simulation of the system depicted in Fig. 4, evidences in light gray the car closer to the viewer, and in darker gray the approaching car in the background.

4 Conclusions and perspectives

The analog VLSI perceptual solutions presented in this paper are related to common *neuromorphic engineering* paradigms [8] [1] but differ from other vision chips in the literature, by the approach followed in their specification. The main innovative features concern: (i) the systematic approach to cooperative computation based on RD paradigms (ii) the architectural specification of an analog perceptual agent –the *Perceptual Engine*– to be used as a powerful building block for early vision perceptual tasks; (iii) the VLSI implementation of such recurrent network as an analog VLSI CMOS circuit; (iv) the overall specification of a microsystem combining sensorial and computational capabilities based on the perceptual engine. The main assets of the Perceptual Engine are its capability of providing rapid "reactions" to sensorial data and its robustness. It performs real-time convolutions of the input image sinked at the nodes of the network with Gabor-like kernels and stable kernel solutions are guaranted by the design sinthesis procedure followed.

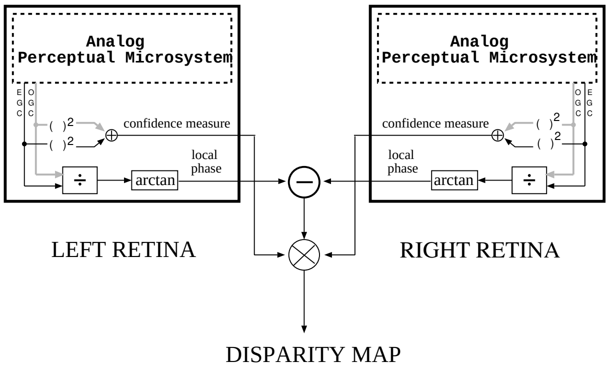
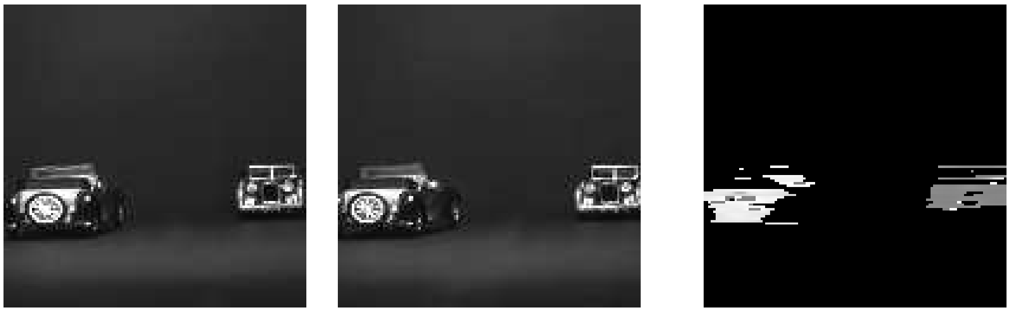


Figure 4: Schematic representation of the architecture of the stereo depth perception subsystem.



left image right image disparity map

Figure 5: Example of depth estimation from a stereo image pair. The disparity map, obtained through a simulation of the system depicted in Figs. 3-4, evidences in light gray the car closer to the viewer, and in darker gray the approaching car in the background.

The perceptual applications presented focus on low-level vision tasks such as stereo disparity estimation and motion analysis. At subsystem level, information processing capabilities at intermediate and high levels could be obtained by our microsystem by adding further processing layers of similar cells. At microsystem level, attractive perspectives concern: (i) the improvement of imager performances, by introducing adaptive contrast gain control circuits; (ii) the extension to 2D processing; and (iii) the implementation of coarse-to-grain (multiscale) strategies [14]. The last one is related to the possibility of adjusting all the kernel's parameters of the Perceptual Engine. To test the feasibility of this approach, a novel chip architecture, based on continuously adjustable transconductors [15], has been recently designed and simulated obtaining results which looks very promising, since. besides the phase, also the frequency and the spatial extension of the Perceptual Engine kernel can be continuously tuned.

Acknowledgments The authors wish to thank Dr. L. Raffo, Ing. G.M. Bo, Ing. M. Confalone, and Ing. M.P. Zizola. This work was partially supported by the *CEC- Esprit BR Project CORMORANT-8503*, by the *University of Genoa Project "Strutture Fisiche per la Percezione e l'Elaborazione", and by* ASI Project "Microsistemi Ottici per la Robotica Spaziale".

References

[1] C. Koch and B. Mathur. Neuromorphic vision chips. *IEEE Spectrum*, 33(5):38–46, 1996.

[2] G.H. Granlund, H. Knutsson, C.-J. Westelius, and J. Wiklund. Issues in robot vision. *Image and Vision Comp.*, 12:131–148, 1994.

[3] V. Virsu, J. Rovamo, P. Laurinen, and R. Nasanen. Temporal contrast sensitivity and cortical magnification. *Vision Res.*, 22:1211–1217, 1982.

[4] S. Okabayashi, M. Furukawa, and T. Hatada. Experimental study on superiority of automotive head-up displays in visual optics. *J. of the Institute of Television Engineers of Japan*, 47:714–721, 1993.

[5] Y. Kuramoto. *Chemical Oscillations, Waves and Turbolences.* Springer, Berlin, 1984.

[6] P.C. Fife. Mathematical aspects of reacting and diffusing systems. *Lecture Notes in Biomathematics*, 28, 1979.

[7] L. Raffo. Analysis and synthesis of resistive networks for distributed visual elaborations. *Electronics Letters*, 32(8):743, April 1996.

[8] C.A. Mead. *Analog VLSI and Neural Systems.* Addison-Wesley, Reading, 1989.

[9] G.M. Bisio, M. Bruccoleri, P. Cusinato, L. Raffo, and S.P. Sabatini. An analog VLSI massively parallel module for low-level cortical processing in machine vision. In *Proc. IEEE-MICRONEURO94 - Torino (Italy)*, 1994.

[10] E.A. Vittoz. Analog VLSI signal processing: Why, where, and how? *J. of VLSI Signal Processing*, 8:27–44, 1994.

[11] J.G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Amer.*, A/2:1160–1169, 1985.

[12] T.D. Sanger. Stereo disparity computation using Gabor filters. *Biol. Cybern.*, 59:405–418, 1988.

[13] I.P. Howard and B.J. Rogers. *Binocular vision and stereopsis.* Oxford University Press, 1995.

[14] H. Wechsler. Multiscale and distributed visual representations and mappings for invariant low-level perception. In H. Wechsler, editor, *Neural Networks for Perception I*, pages 462–476. Academic Press, 1992.

[15] F. Bruccoleri, L. Raffo, S.P. Sabatini, and G.M. Bisio. A tunable perceptual microsystem for stereo depth estimation. In *2nd IEEE-CAS Region 8 Workshop on Analog and Mixed IC Design*, Baveno, Italy, September 1997.