Visual Perception Microsystems Based on Distributed Analog VLSI Processing

L. Raffo[b], S.P. Sabatini[a], G.M. Bo[a], and G.M. Bisio[a]
[a]PSPC Research Group-DIBE University of Genoa
Via all'Opera Pia 11A, 16145 Genova, ITALY e_mail: silvio@DIBE.unige.it
[b]DIEE University of Cagliari
Piazza d'Armi, 09123 Cagliari, ITALY e_mail: luigi@DIEE.unica.it

Abstract

The paper discusses the potentialities and feasibility of a vision system to be implemented as a neuromorphic microsystem, thus satisfying the physical and economical constraints of low-end components. The paper presents methods and approaches to the design of micropower collective computational systems based on an appropriate cooperative organization of building blocks. First, to characterize the essence of recurrent lattice networks, 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 edge detection, texture analysis, stereo disparity estimation, and motion analysis, is considered.

1 Introduction

In recent years, many new approaches have been taken to perform tasks based on the responses to visual information. They are motivated by a need to *react* to visual stimuli in appropriate ways, resulting in a coupling between behavior and perception. In general, such systems could be though as *behavioral systems* able to provide a *reflex* as a shortcut to action. Though the feasibility of a machine vision approach to early-vision tasks has been proven, their effective and low-cost implementation have not yet. The efficacy of the algorithm, indeed, has usually to cope with the efficiency of hosting computational architecture. Unfortunately, despite continuous advances in microelectronics technology, hardware simulation of machine vision algorithms still remain well below the performance level expected. To bridge this gap, 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) [Kobayashi *et al.*1991] [Arreguit and Vittoz1994] [Wyatt *et al.*1992a] [Koch and Mathur1996].

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, indeed, is really distributed, being mapped directly in the electrical variables, and computation is carried out massively in parallel with high efficiency and speed. Information processing occurs at 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 [Arreguit *et al.*1993] [Andreou *et al.*1991] [Andreou and Boahen1996] [Koch and Mathur1996]. Hence, rather than information processing systems, it should be more appropriate to refer to perceptual microsystems as "information captors", where information is intended to be the structural organization of sensorial signals.

Effective collective computation (small area and low-power consumption) on real-time signals is only possible when taking full advantage of the CMOS technology at transistor and/or circuit levels. Several examples are present in the literature illustrating how powerful this approach can be. Among these contributions we can distinguish two schools of thought. The first one, problem oriented, starts from a computational approach to visual tasks, formulated as variational problems. Analog VLSI circuits are considered as "analog computers" on which to map the related Euler-Lagrange equations [Poggio1985] [Koch *et al.*1986] [Kobayashi *et al.*1993]. The second one, stressing the similarities between the physics of analog CMOS and brain circuits, has accomplished the goal of reproducing the various blocks of first stages in the visual pathway: photoreceptors, retinal cells, and optical nerve. Applications of these blocks have been considered in relation to contrast adaptation, stereo-correspondence, motion, etc. [Mead1989] [Delbrück1993] [Wyatt *et al.*1992b] [Shi *et al.*1993].

In this scenario, guided by the neurophysiological findings on the filtering properties of visual neurons, our research intends to demonstrate how a variety of computational visual tasks can be mapped on the same "medium" which can be implemented efficiently through linear CMOS circuits. Early vision perceptual processes can be interpreted as a "measuring" operation on the visual signal [Adelson and Bergen1991]. 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. A proper formalization of these concepts is presented in Section 2 introducing lattice networks and analyzing their behavior in the Z-transform domain. In this way, it can be evidenced how each type of lattice network can be sustained by a unified entity, or computational *engine*, on which the various perceptual algorithms can be mapped by acting properly on the pattern of afferent connections. The functionality of 2nd-order lattice networks is discussed in detail and their VLSI implementation is verified (see Section 3). The use of this basic cooperative block in relevant perceptual problems such as edge detection, texture analysis, stereo disparity estimation, or motion analysis is discussed (see Section 4).

The activity presented in this work has been mainly conducted in the framework of the ESPRIT Basic Research Action project 8503, Cormorant. The main objective is to abstract from cortical processing to analogue hardware solutions for effective perceptual modules. In particular, the project has been tackling the problem of dynamic depth perception in order to support navigation and object identification.

2 Lattice models for early vision

A great deal of visual tasks is solved by animals and humans in a preattentive phase. In this phase, vast arrays of intercommunicating, identical processes carry out early vision tasks such as perceiving edges by texture differences, depth mapping, computing optical flow, recovering local surface structure, segmenting images, etc. These visual tasks can be formalized in terms of spatially continuous linear operators specified in differential and/or integral forms [Freeman and Adelson1991] [Poggio *et al.*1985] [Koenderink1988]. The discrete forms of these 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:

$$\sum_{k,l} a_{kl} e(m-k, n-l) \sum_{k,l} b_{kl} s(m-k, n-l) \qquad (1)$$

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. 1 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 $h(\cdot,\cdot)$:

$$e(m,n) = \sum_k \sum_l h(m-k, n-l) s(k,l). \qquad (2)$$

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 $H(z_1, z_2)$ of the impulse response:

$$H(z_1, z_2) = \frac{\sum_{k=-I_m}^{I_m} \sum_{l=-I_n}^{I_n} b_{kl} z_1^{-k} z_2^{-l}}{\sum_{k=-O_m}^{O_m} \sum_{l=-O_n}^{O_n} a_{kl} z_1^{-k} z_2^{-l}}. \qquad (3)$$

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

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

Under these conditions, the 2D problem can be recast as the cascade of two 1D problems. As shown in the following, this assumption does not impact adversely on the solutions of the problems we are addressing.

We intend to investigate the functionality of 1-D lattice networks with respect to the degree of locality of interactions both excitatory and inhibitory. In order to better single out the essence of cooperative computation, we consider only the recurrent contributions, and assume the stimulus to be applied with just a one-to-one feed-forward mapping. This will allow us to introduce the concept of "perceptual engine", thus pointing out how the same computational model can be used as a basic building block for a whole variety of filtering tasks.

In the following, we can assume that the perceptual engine is symmetric since, afterwards, we can regain generality of forms for the kernel acting upon the b's coefficients. Given the difference equation describing the nearest neighbor cooperation among nodes:

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

one derives the Z-domain transfer function and obtains the spatial impulse response [Raffo1996]:

$$h^{(1)}(n)=C^{(1)}\mathrm{e}^{-\lambda|n|} \qquad \lambda>0 \tag{6}$$

where $b_0=C^{(1)}(1-\mathrm{e}^{-2\lambda})$, $a_{-1}=a_1=-\mathrm{e}^{-\lambda}$, and $a_0=\mathrm{e}^{-2\lambda}+1$.

Given the difference equation describing the nearest and next nearest neighbor cooperation among nodes:

$$\begin{aligned} a_{-2}e(n-2)+a_{-1}e(n-1)+a_0e(n) & + \\ +a_1e(n+1)+a_2e(n+2) &= b_0s(n) \end{aligned} \tag{7}$$

one derives the Z-domain transfer function and obtains the spatial impulse response [Raffo1996]:

$$\begin{aligned} h^{(2)}(n) &= C^{(2)}\mathrm{e}^{-\lambda|n|}((1+\mathrm{e}^{-2\lambda})\sin(\omega_0)\cos(\omega_0 n) \\ &+ (1-\mathrm{e}^{-2\lambda})\cos(\omega_0)\sin(\omega_0|n|)) \end{aligned} \tag{8}$$

where $b_0=C^{(2)}[(1-\mathrm{e}^{-4\lambda})(1+\mathrm{e}^{-2\lambda})\sin(\omega_0)-4\mathrm{e}^{-2\lambda}(1-\mathrm{e}^{-2\lambda})\sin(\omega_0)\cos^2(\omega_0)](1-\mathrm{e}^{-4\lambda})$, $a_{-2}=a_2=\mathrm{e}^{-2\lambda}$, $a_{-1}=a_1=-2\mathrm{e}^{-\lambda}(1+\mathrm{e}^{-2\lambda})\cos\omega_0$, $a_0=1+4\mathrm{e}^{-2\lambda}\cos^2\omega_0+\mathrm{e}^{-4\lambda}$. 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. These functions ($h^{(1)}$ and $h^{(2)}$) can be interpreted as the "perceptual engine" of our systems, providing computational primitives that can be composed to obtain more powerful image descriptors.

In general, we can state that any convolution kernel, whose two-sided Z transform has at the denominator a fourth-order polynomial with symmetric coefficients, can be obtained by a weighted local sum over the output of the 1-D 2nd-order cooperative network. Typical convolution kernels useful for visual tasks (difference of Gaussians (DOG) [Marr1982], difference of offset Gaussians (DOOG) [Parker and Hawken1988] [Malik and Perona1990], Gabor functions [Daugman1985] [Porat and Zeevi1988]) can be implemented by a 2nd order network [Raffo1996]. 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:

$$\begin{aligned} g(n) &= \alpha h(n-1)+\beta h(n)+\gamma h(n+1) \\ &= C\mathrm{e}^{(-\lambda|n|)}\cos(\omega_0 n+\varphi). \end{aligned} \tag{9}$$

where $\alpha=C\mathrm{e}^{-\lambda}(\mathrm{e}^{-2\lambda}-1)\cos(\omega_0-\varphi)$ $\beta=C(1-\mathrm{e}^{-4\lambda})\cos(\omega_0)$ $\gamma=C\mathrm{e}^{-\lambda}(\mathrm{e}^{-2\lambda}-1)\cos(\omega_0+\varphi)$. 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.

Section 4 presents an overview of this approach for various machine vision tasks, evidencing how the functionality of the perceptual engines can provide powerful computational resources.

3 Distributed computation on analog physical systems

Spatial filters can be implemented by using linear resistive networks with positive and negative resistors [Mead1989] [Kobayashi *et al.*1991] [Arreguit and Vittoz1994] [Raffo1995b], or by current generators driven by the voltage at other nodes (VCCS) [Bisio *et al.*1994]. 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 architectureof one-way-interacting elements implemented as CCCS. One advantage of this interpretation shows up at transistor level, where it is shown how to implement these interactions with "high" precision through current mirrors.

Circuit level Let us 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 current controlled current sources (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 [Vittoz1994], thus eliminating the dependence on the performances of single devices.

Assuming $a_0\neq 0$, $a_1=a_{-1}$ and, $a_2=a_{-2}$, thus we can define $G=-a_1/a_0$ and $K=a_2/a_0$ and rewrite Eq. 7 as:

$$\begin{aligned} I_e(n) &= I_s(n)+GI_e(n-1)+GI_e(n+1)- \\ &- KI_e(n-2)-KI_e(n+2) \end{aligned} \tag{10}$$

$$e(n)=\frac{I_e(n)}{I_{norm}}, \qquad \frac{b_0}{a_0}s(n)=\frac{I_s(n)}{I_{norm}}$$

where $I_s(n)$ is the normalized input current (i.e., stimulus) and $I_e(n)$ is the normalized output current (i.e. excitation); the coefficients G and K are always positive. Eq. 10 can be interpreted in architectural terms as an array of circuital computational nodes (see Fig. 1), with local interconnections both at the nearest and the next nearest neighbors. Each computational node has two

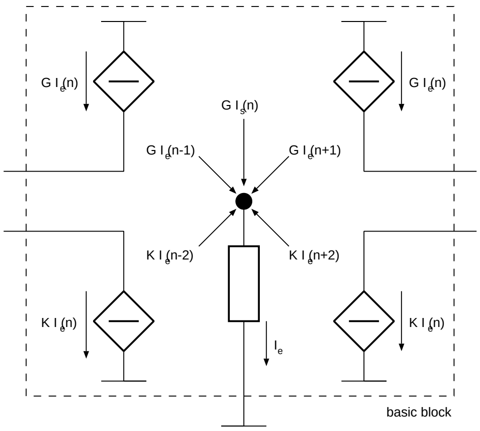


Figure 1: The node basic block and its internal circuital structure: the black dot represents the node which is stimulated by the photocurrent $I_s(n)$ and receives inhibitory (K) and excitatory (G) contributions from neighboring nodes. The sum of the current contributions from other nodes is performed by Kirchhoff current law at the node. The resulting excitation current, $I_e(n)$, acts, in turn, on neighboring nodes controlling the currents of four current generators.

output currents $GI_e(n)$ toward the nearest nodes, two (negative) output currents $KI_e(n)$ toward the next nearest nodes, and receives the corresponding contributions from its neighbors, besides its input excitation $I_s(n)$. The circuit representation of a node is based on the use of CCCSs with the desired current gain G and K (see Fig. 1(b)).

Transistor level As shown in Fig. 2, the CCCSs are implemented by using standard current mirrors which provide, when the transistors are operating in the saturation region, a weighted copy of their input currents according to the W/L ratios of the two transistors. To improve the matching between devices, a current gain p/q, different from one, is best obtained by using two sets of identical transistors connected in parallel (i.e. q for T1 and a proper p_g for T2 and p_k for T3 and T4). The two CCCSs with gain K are obtained through the current mirror transistors T1-T3(T4); the two CCCSs with gain G are obtained through the current mirror transistors T1-T2-T5-T6(T7). It is worth noting that in this

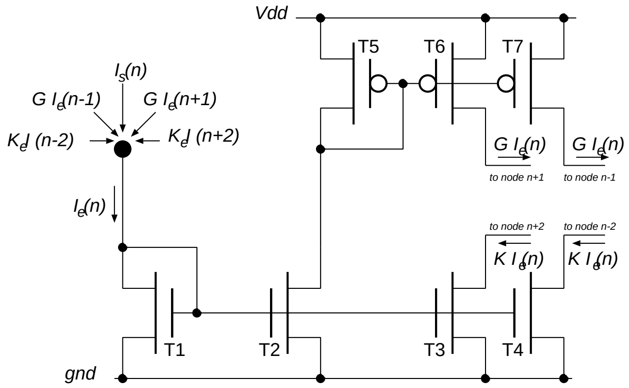


Figure 2: Description of the node basic block at transistor level. $I_e(n)$ is copied (weighted according to G and K values) to provide excitation or inhibition to the neighboring nodes.

way the values of the gains K and G are restricted to rational numbers, in relation to all possible combinations of the set of identical transistors connected in parallel in each branch of the current mirrors; these combinations, selectable by proper switches, are characterized by the same level of precision.

If one considers that a current mirror is a cell in which the input and output currents are monodirectional, whereas the signals of the filter (mapped on currents) can be both positive and negative, one concludes that the circuit implementation of each generic node should be composed of two complementary parts, one for positive input currents (i.e., the current is injected into the node), and one for negative input currents (when the current is sourced from the node). The circuit solution for the negative input current is the same of the one shown in Fig. 2, replacing the NMOS with PMOS transistors and vice versa. In the complete schematic of a node two transistors are introduced to cancel the current flowing through VDD and GND when the input current I_{in} is null. In this way only one of the two complementary parts of the cell is active when an input current is injected or sourced to the cell.

The present implementation of the chip (see Fig. 3) is characterized by $q = 7$, $p_g = 1$, and $p_k = 2$. In this way, the gain of the CCCSs are $G = 1/7$ and $K = 2/7$ which correspond to have $\lambda = 0.5713$ and $\omega_0 = 1.46354$.

As discussed, by combining the response of neighboring nodes it is possible to obtain a Gabor-like function of proper phase (decay rate and frequency are hard-wired in the underlying perceptual engine). Figure 4 compares measured and expected impulse response for three different values of the phase: 0, $\pi/2$ and $\pi/4$. These results

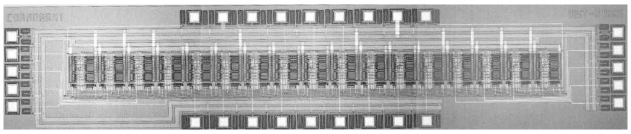


Figure 3: Microphotograph of a 1-D 2nd order network chip

evidence the feasibility of the perceptual engine approach to the generation of more complex functionalities.

4 Potential applications

Early vision tasks concern the analysis of visual information by the decomposition of the input image into coherent components that can be independently characterized [Marr1982]. Thus, 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 [Daugman1985] [Porat and Zeevi1988] [Fogel and Sagi1989] [Chang and Chatterjee1993].

In this section, we'll examine the application capabilities of our perceptual microsystem as a front-end module in a machine vision system. The perceptual engine can be used as an elementary operator on its own, or as a building block to obtain more elaborate visual operators.

Edge detection Edge detection is a key first step in early vision. The problem of measuring sharp brightness changes in the image can be efficiently solved by the application of an even/odd pair of Gabor functions [Burr and Morrone1992]. As shown in Fig. 4a and b, a convolution operation with a pair of 1-D Gabor-like kernels can be obtained by feeding a line of the image to the perceptual engine and by combining the outputs at the nodes in a symmetric and antisymmetric fashion. If the application requires the use of 2-D convolution operators with $x - y$ separable kernels, it is possible to get this result by feeding the output of a rank of 1D 2nd-order networks lying in a specific axis to an orthogonal rank of similar networks (see Fig. 5).

Texture analysis The same approach can be pursued to generate 2-D Gabor-like kernels capable to detect both the positions and the orientations of the edges (see Fig. 6). Systems based on a small set of such filters, differently oriented, can be used to evidence texture differences in an image by using the responses of networks selective to various orientations (i.e., lying along different axes) [Turner1986] [Malik and Perona1990] [Bergen and Landy1991] [Jain and Farrokhnia1991] [Manjunath and Chellappa1993] [Indiveri *et al.*1995]: the presence of repeated oriented elements shows up as the strongest output for the networks of the corresponding orientations.

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. Hence, a joint representation of the information flows coming from the two eyes is a definitive need for stereoscopic depth perception. Such a joint representation can be related to the spatial properties of binocular filters tuned to specific disparity values [Sanger1988]. This problem has an elegant formulation in the frequency domain, in relation to the so-called phase-based matching methods [Sanger1988]. 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. Exploiting the properties of the perceptual engine one can change the phase of the resulting Gabor filter in order to find the local phase by an adaptive procedure [Raffo1995a]. We applied the algorithm to a synthetic stereo pair; in Fig.7 three boxes are located at different distances from the viewer: the larger one, on the right, is at the same distance as the background. The resulting depth map is depicted in Fig.7(c): bright grey corresponds to objects closer to the viewer.

Motion analysis Several approaches to motion analysis rely upon correspondence measures in the spatio-temporal domains of a visual scene. Among the others, the so-called "energy" or "spatio-temporal" models [Adelson and Bergen1985] are based on a linear filtering operation with kernels with a specific orientation in the spatio-temporal plane which gives them a specific direction sensitivity. A set of such filters are often used as a first stage of motion algorithms in order to extract information about local velocity [Nowlan and Sejnowski1994].

Adelson and Bergen [Adelson and Bergen1985] demonstrated that spatio-temporal oriented filters can be constructed from pairs of separable (i.e., not oriented) filters. According to the method of Adelson and Bergen, filters tuned to specific directions can be obtained through a temporally modulated combination of perceptual engines (see Fig. 8). Specifically, by taking appropriate sums and differences of the outputs of a 2nd-order lattice network, it is possible to construct two pairs of oriented filters to be used, as shown at the bottom of Fig. 8 to measure and discriminate the energy of leftward and rightward motion:

$$\begin{aligned}
g_{even,3}(n,t) &= \alpha_{even} f_3(t)h(n-1) + \beta_{even} f_3(t)h(n) + \\
&+ \gamma_{even} f_3(t)h(n+1) \\
g_{odd,3}(n,t) &= \alpha_{odd} f_3(t)h(n-1) + \gamma_{odd} f_3(t)h(n+1) \qquad (11) \\
g_{even,5}(n,t) &= \alpha_{even} f_5(t)h(n-1) + \beta_{even} f_5(t)h(n) + \\
&+ \gamma_{even} f_5(t)h(n+1) \qquad (12) \\
g_{odd,5}(n,t) &= \alpha_{odd} f_5(t)h(n-1) + \gamma_{odd} f_5(t)h(n+1)
\end{aligned}$$

where

$$f_k(t) = (\omega_t t)^k \mathrm{e}^{-\omega_t t} \left[\frac{1}{k!} - \frac{(\omega_t t)^2}{(k+2)!} \right]$$

Following a similar procedure, if needed, similar filters for upward and downward motion can be straightforwardly obtained.

5 Discussion and conclusions

The approach followed in this paper can be related to the research path followed by Carver Mead and his colleagues [Mead1989] towards efficient silicon implementations of visual tasks. Simple cooperative networks are seen both as a medium whose structure is in many ways similar to neural structures [Sivilotti *et al.*1987] [Mead and Mahowald1988], and as the circuit analogue of the discretized forms of machine vision algorithms [Koch1989] [Koch *et al.*1988]. The main innovative feature of our approach concerns the architectural specification of an analog perceptual agent –the *perceptual engine*– to be used as a powerful building block for early vision perceptual tasks. The computational capabilities of the perceptual engine is based on the collective computation of a lattice network where the excitation value of a node is related to the excitations of its neighbors. The overall computation is represented by the pattern of activity at the nodes of the networks.

At architectural level, it does suit to think of a node as an "interacting site" made of the recipient node together with its afferent interconnections. This description leads to a direct VLSI implementation in which both the excitation and the stimuli are mapped on current. The node basic block collects all positive and negative contributions from afferent nodes and generates currents proportional to its excitation as input to neighboring nodes. To achieve the necessary precision the gain of its current mirror is determined by the number of identical transistors in parallel for each branch. In this way the gain (i.e. frequency and decay length of perceptual engines) can be selected in a wide set of discrete values with the same "high" precision.

At implementation level the soundness of the design approach is verified both for the perceptual engine response and for its composition to achieve more complex functionalities.

The perceptual applications presented focus on low-level vision tasks such as edge detection, extraction of stereo disparity, and analysis of motion. In perspective, information processing capabilities at intermediate and high levels could be obtained by our microsystem by adding further processing layers of similar cells. The process can be recursively applied to consider operations like segmentation, recovery of surface structures, or recognition of simple objects.

The studies on the structure of mammalian visual systems could further suggest architectural models based on simple computational units, with dense local interconnections, apt to perform complex visual tasks in machine vision also for real world scenes, and possibly to overcome the limitations of many AI algorithms strongly oriented toward specific applications. Specifically, if we deal with analog computation, where the physical process *is* the computation, special attention should be paid to the physical structure of the neocortex since it is tightly related to its computational properties. The specific organization of the neocortex [Douglas and Martin1992], its economic wiring [Mitchison1992], and the relations to the properties of cortical neurons justify, indeed, the efforts to understand in what sense cortical circuits transform their biology into computation.

Acknowledgments The authors wish to thank Prof. G. Soncini and Ing. G.F. Dalla Betta for the support in the fabrication of the chip, Dr. A. Bonfiglio for the testing of the chip, and Ing. F. Solari for contributions to the application examples. This work was partially supported by the *CEC- Esprit BR Project CORMORANT-8503*, and by the *University of Genoa Project "Strutture Fisiche per la Percezione e l'Elaborazione"*.

References

[Adelson and Bergen1985] E.H. Adelson and J.R. Bergen. Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Amer.*, 2:284–321, 1985.

[Adelson and Bergen1991] E.H. Adelson and J.R. Bergen. The plenoptic function and the elements of early vision. In M.S. Landy and J.A. Movshon, editors, *Computational Models of Visual Processing*, pages 3–20. MIT Press, 1991.

[Andreou and Boahen1996] A.G. Andreou and K.A. Boahen. Translinear circuits in subthreshold MOS. *Analog Integrated Circuits and Signal Processing*, 9:141–166, 1996.

[Andreou *et al.*1991] A.G. Andreou, K.A. Boahen, P.O. Pouliquen, A. Pavasović, R.E. Jenkins, and K. Strohbehn. Current mode subthreshold MOS circuits for analog VLSI neural systems. *IEEE Trans. Neural Net.*, 2:205–213, 1991.

[Arreguit and Vittoz1994] X. Arreguit and E.A. Vittoz. Perception systems implemented in analog VLSI for real-time applications. In *Proc. of the PerAc'94 Conference: From Perception to Action*, Lausanne, Sept.5-9, 1994, 1994.

[Arreguit *et al.*1993] X. Arreguit, E.A. Vittoz, F.A. Schaik van, and A. Mortara. Analog implementation of low-level vision systems. In *Proc. ECCTD'93*, Davos, Aug.30-Sept.3, 1993, 1993.

[Bergen and Landy1991] J.R. Bergen and M.S. Landy. Computational modeling of visual texture segregation. In M.S. Landy and J.A. Movshon, editors, *Computational Models of Visual Processing*, pages 253–271. MIT Press, 1991.

[Bisio *et al.*1994] 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)*, pages 207–211, 1994.

[Burr and Morrone1992] D.C. Burr and M.C. Morrone. A nonlinear model of feature detection. In *Nonlinear Vision: Determination of Neural Receptive Fields, Functions, and Networks*, pages 309–328. CRC press, 1992.

[Chang and Chatterjee1993] C. Chang and S. Chatterjee. Ranging through Gabor logons - A consistent, hierarchical approach. *IEEE Trans. Neural Net.*, 4(5):827–843, 1993.

[Daugman1985] 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.

[Delbrück1993] T. Delbrück. Silicon retina with correlation-based velocity-tuned pixels. *IEEE Trans. Neural Net.*, 4:529–541, 1993.

[Douglas and Martin1992] R.J. Douglas and K.A.C. Martin. A functional microcircuit for cat visual cortex. *J. Physiol.*, 440:735–769, 1992.

[Fogel and Sagi1989] I. Fogel and D. Sagi. Gabor filters as texture discriminator. *Biol. Cybern.*, 61:103–113, 1989.

[Freeman and Adelson1991] W.T. Freeman and E.H. Adelson. The design and use of steerable filters. *IEEE Trans. Pattern Anal. Mach. Intell.*, 13(9):891–906, 1991.

[Indiveri *et al.*1995] G. Indiveri, L. Raffo, S.P. Sabatini, and G.M. Bisio. A neuromorphic architecture for cortical multi-layer integration of early visual tasks. *Machine Vision and Applications*, 8:305–314, 1995.

[Jain and Farrokhnia1991] A.K. Jain and F. Farrokhnia. Unsupervised texture segmentation using Gabor filters. *Pattern Recognition*, 24:1167–1186, 1991.

[Kobayashi *et al.*1991] H. Kobayashi, J.L. White, and A.A. Abidi. An active resistor network for Gaussian filtering of images. *IEEE J. on Solid State Circuits*, 26:738–748, May 1991.

[Kobayashi *et al.*1993] H. Kobayashi, T. Matsumoto, T. Yagi, and T. Shimmi. Image processing regularization filters on layered architecture. *Neural Networks*, 6:327–350, 1993.

[Koch and Mathur1996] C. Koch and B. Mathur. Neuromorphic vision chips. *IEEE Spectrum*, 33(5):38–46, 1996.

[Koch *et al.*1986] C. Koch, J. Marroquin, and A. Yuille. Analog "neuronal" networks in early vision. *Proc. Natl. Acad. Sci.*, 83:4263–4267, 1986.

[Koch *et al.*1988] C. Koch, J. Luo, and Mead. C. Computing motion using resistive networks. In D.Z. Anderson, editor, *Neural Information Processing Systems*, pages 422–431. 1988.

[Koch1989] C. Koch. Seeing chips: Analog VLSI circuits for computer vision. *Neural Comp.*, 1:184–200, 1989.

[Koenderink1988] J.J. Koenderink. Scale-time. *Biol. Cybern.*, 58:159–162, 1988.

[Malik and Perona1990] J. Malik and P. Perona. Preattentive texture discrimination with early vision mechanisms. *J. Opt. Soc. Amer.*, 7:923–932, 1990.

[Manjunath and Chellappa1993] B.S. Manjunath and R. Chellappa. A unified approach to boundary perception: Edges, textures and illusory contours. *IEEE Trans. Neural Net.*, 4:96–107, 1993.

[Marr1982] D. Marr. *Vision*. New York: Freeman, 1982.

[Mead and Mahowald1988] C.A. Mead and M.A. Mahowald. A silicon model of early visual processing. *Neural Networks*, 1:91–97, 1988.

[Mead1989] C.A. Mead. *Analog VLSI and Neural Systems.* Addison-Wesley, Reading, 1989.

[Mitchison1992] G. Mitchison. Axonal trees and cortical architecture. *Trends in Neurosci.*, 15:122–126, 1992.

[Nowlan and Sejnowski1994] S.J. Nowlan and T.J. Sejnowski. Filter selection model for motion segmentation and velocity integration. *J. Opt. Soc. Amer.*, 11:3177–3200, 1994.

[Parker and Hawken1988] A. Parker and M.J. Hawken. Two dimensional spatial structure of receptive fields in monkey striate cortex. *J. Opt. Soc. Amer.*, 5(4):598–605, 1988.

[Poggio *et al.*1985] T. Poggio, V. Torre, and C. Koch. Computational vision and regularization theory. *Nature*, 317:638–643, 1985.

[Poggio1985] T. Poggio. Early vision: from computational structure to algorithms and parallel hardware. *Computer Vision, Graphics, and Image Processing*, 31:139–155, 1985.

[Porat and Zeevi1988] M. Porat and Y.Y. Zeevi. The generalized Gabor scheme of image representation in biological and machine vision. *IEEE Trans. Pattern Anal. Mach. Intell.*, 10(4):452–467, 1988.

[Raffo1995a] L. Raffo. Adaptive resistive network for stereo depth estimation. *Electronics Letters*, 31(22):1909–1910, 1995.

[Raffo1995b] L. Raffo. Resistive network implementing maps of Gabor functions of any phase. *Electronics Letters*, 31(22):1913–1914, 1995.

[Raffo1996] L. Raffo. Analysis and synthesis of resistive networks for distributed visual elaborations. *Electronics Letters*, 32(8):743, April 1996.

[Sanger1988] T.D. Sanger. Stereo disparity computation using Gabor filters. *Biol. Cybern.*, 59:405–418, 1988.

[Shi *et al.*1993] B.E. Shi, T. Roska, and L.O. Chua. Design of linear cellular neural networks for motion sensitive filtering. *IEEE Trans. on Circuit and Systems: I. Fundamental Theory and Applications*, 40:320–331, 1993.

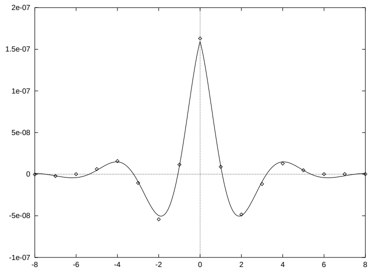
[Sivilotti *et al.*1987] M.A. Sivilotti, M.A. Mahowald, and C.A. Mead. Real-time visual computations using analog CMOS processing arrays. In *Advanced Research in VLSI:Proc. Stanford Conference 1987*, pages 295–312, 1987.

[Turner1986] M.R. Turner. Texture discrimination by Gabor functions. *Biol. Cybern.*, 55:71–82, 1986.

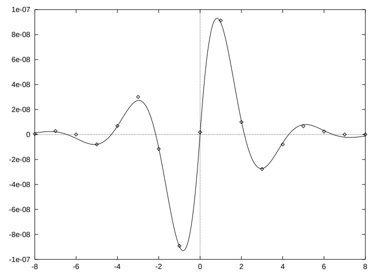
[Vittoz1994] E.A. Vittoz. Analog VLSI signal processing: Why, where, and how? *J. of VLSI Signal Processing*, 8:27–44, 1994.

[Wyatt *et al.*1992a] J.L. Wyatt, C. Keast, M. Seidel, D. Standley, B. Horn, T. Knight, C. Sodini, H. Lee, and T. Poggio. Analog VLSI systems for image acquisition and fast early vision processing. *Int. J. of Comp. Vision*, 8:217–230, 1992.

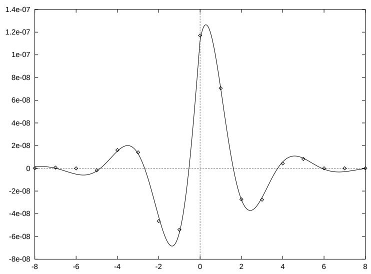
[Wyatt *et al.*1992b] J.L. Wyatt, C. Keast, M. Seidel, D. Standley, B. Horn, Knight T., C. Sodini, H.-S. Lee, and T. Poggio. Analog VLSI systems for image acquisition and fast early vision processing. *Int. J. of Computer Vision*, 8:217–230, 1992.



(a)



(b)



(c)

Figure 4: Comparisons between measured (dots) and theoretical (continuous lines) Gabor-like impulse responses of phases 0 (a), $\pi/2$ (b) and $\pi/4$ (c).

(a)

(b)

Figure 5: (a) A test image and its edge detection (b) using two orthogonal 17-node 1-D kernels in cascade.

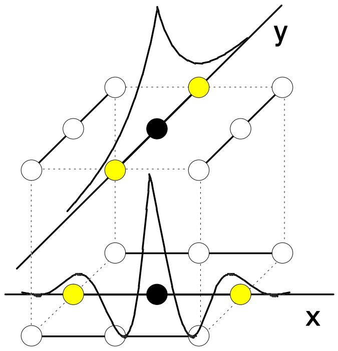
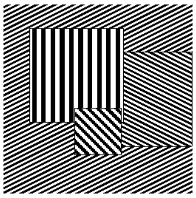
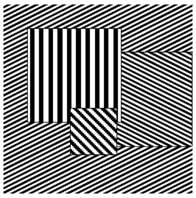


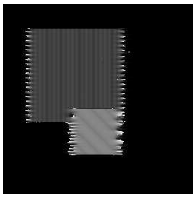
Figure 6: Generation of a 2D Gabor-like filter by the cascade of two 1D orthogonal lattice networks.



(a)



(b)



(c)

Figure 7: A pair of artificial stereo images (a) and (b), and the result of the algorithm (c): bright grey corresponds to objects closer to the viewer.

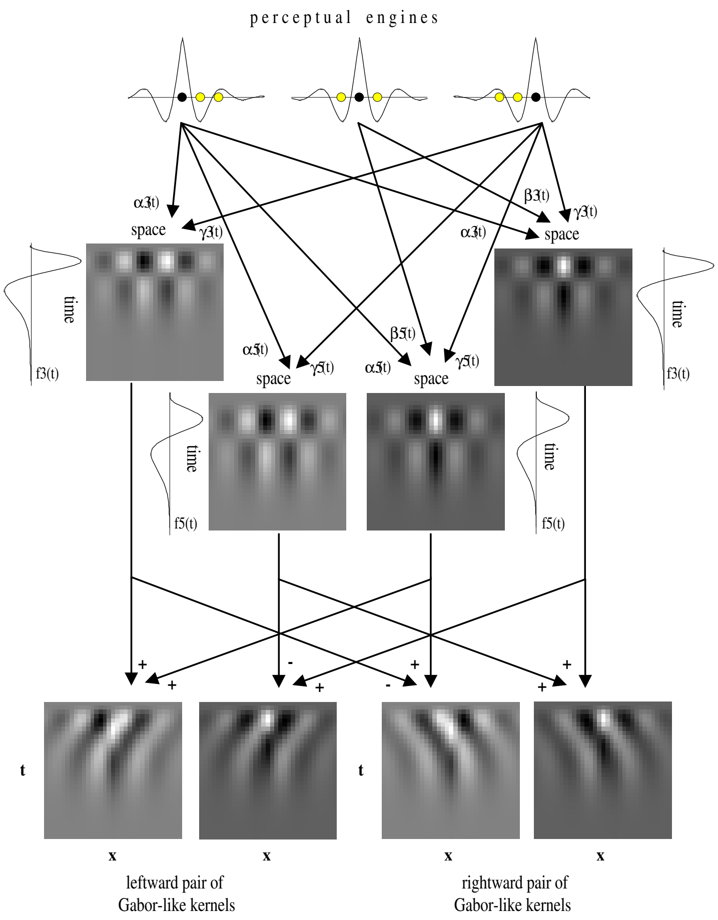


Figure 8: Generation of motion sensitive filters through a temporal modulated local combination of perceptual engines. The resulting filters, shown at the bottom of the figure, are a pair of leftward, and a pair of rightward selective filters; members of a pair are approximately in quadrature. It should be observed that the temporal dependence has been introduced here on a merely descriptive basis. Adapted from (Adelson and Bergen, 1985).