# Implementing early vision algorithms in analog hardware: an overview

Christof Koch

Christof Koch, "Implementing early vision algorithms in analog hardware: an overview," Proc. SPIE 1473, Visual Information Processing: From Neurons to Chips, (9 July 1991); doi: 10.1117/12.45546

SPIE.

Event: Orlando '91, 1991, Orlando, FL, United States

**Invited Paper**

# Implementing early vision algorithms in analog hardware
## An overview

Christof Koch

California Institute of Technology, Computation and Neural Systems 216-76
Pasadena, California 91125

## ABSTRACT

In the last ten years, significant progress has been made in understanding the first steps in visual processing. Thus, a large number of algorithms exist that locate edges, compute disparities, estimate motion fields and find discontinuities in depth, motion, color and intensity. However, the application of these algorithms to real-life vision problems has been less successful, mainly because the associated computational cost prevents real-time machine vision implementations on anything but large-scale expensive digital computers. We here review the use of analog, special-purpose vision hardware, integrating image acquisition with early vision algorithms on a single VLSI chip. Such circuits have been designed and successfully tested for edge detection, surface interpolation, computing optical flow and sensor fusion. Thus, it appears that real-time, small, power-lean and robust analog computers are making a limited comeback in the form of highly dedicated, smart vision chips.

## 1. INTRODUCTION[1]

Vision is simple! We open our eyes and, instantly, the world surrounding us is perceived in all its splendor: "Various books in their colorful dust-jackets clutter up my desk and I can see people move beyond the window." Yet Artificial Intelligence has been trying with limited success for over 20 years to endow machines with similar abilities. A large van, filled with computers and driving unguided at a mile per hour across gently sloping hills in Colorado is the most we have accomplished so far. On the other hand, computers can play a decent game of chess or prove simple mathematical theorems! It is ironic that we are unable to reproduce perceptual abilities which we share with most animals while some of the features distinguishing us from even our closest cousins, chimpanzees, can be carried out by machines. Vision is difficult!

In the last ten years, significant progress has been made in understanding the first steps in visual processing. Thus, a large number of well-studied algorithms exist that locate edges, compute disparities along these edges or over areas, estimate motion fields and find discontinuities in depth, motion, color and texture (for an overview of current research see Horn[1] and the special issue of the *IEEE Proceedings* in August of 1988 on computer vision). Two key problems remain. One is the integration of information from different modalities, i.e., how can disparity obtained from binocular stereo be combined with the depth information contained within the motion field?[2,3] Fusion of information is expected to greatly increase the robustness and fault tolerance of current vision systems as it is most likely the key towards fully understanding vision in biological systems. The second, more immediate, problem is the fact that vision is very expensive in terms of computer cycles. Thus, one second worth of black and white TV adds up to approximately 64 million bits which need be transmitted and further processed (500 by 500 picture elements (pixels) sampled about 32 times per second at 8 bit per pixel). Since early vision algorithms are formulated as relaxation algorithms which need to be executed many hundreds of times before convergence, even supercomputers take their time. For instance, the 65,536 processor Connection Machine at Thinking Machines Corporation,[4] with a machine architecture optimal from the point of view of processing two-dimensional images, still requires several seconds *per image* to compute depth from two displaced images.[5] Performance on 68030, 80386 or RISC microprocessor based workstations is still slower.

Animals, of course, devote a large fraction of their nervous system to vision. Thus, about 270,000 out of 340,000 neurons in the house fly *Musca domestica* are considered to be "visual" neurons,[8] while a third of the human cerebral

[1] This article has been adapted from a previously published chapter.[6]

 0-8194-0582-5/91/$4.00

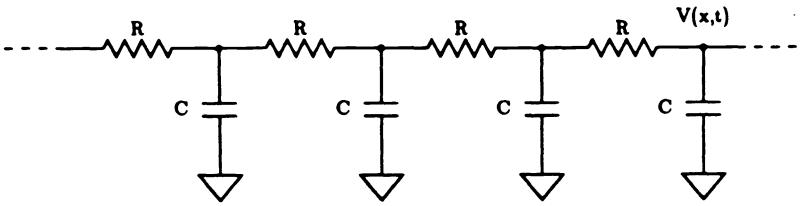


Fig. 1. One-dimensional lumped-element resistor/capacitor transmission line. The incoming light intensity is converted into the initial voltage distribution $V(x,0)$. The voltage $V(x,t)$ along the line is given by the convolution of $V(x,0)$ with a Gaussian of variance $\sigma^2 = t/(RC)$.[7] The voltage at time $t$ then corrresponds to the image intensity convolved with a Gaussian of variance $\sigma^2$. From Koch.[6]

cortex is given over to the computations underlying the perception of depth, color, motion, recognition etc. One way for technology to bypass the computational bottleneck is to likewise construct special-purpose vision hardware. Thus, in 1978 researchers at Hughes Research Laboratories[9] used three 1-D strips of charge-coupled devices (CCD) clocked appropriately, to implement very local convolution operations. They demonstrated blurring and simple edge enhancement algorithms operating at about 10 $kHz$ with a 4-bit intensity resolution.

Today, commercial vendors such as *Imaging Technology* or *Datacube*, offer powerful and programmable digital systems on the open market for a few thousand dollars. Why, however, execute vision algorithms on digital machines when the signals themselves are analog? Why not exploit the physics of circuits to build very compact, analog special-purpose vision systems? This *smart sensor paradigm*, in which as much as possible of the signal processing is incorporated into the sensor and its associated circuitry in order to reduce transmission bandwidth and subsequent stages of computation, is starting to emerge as a possible competitor to more general-purpose digital vision machines.

## 2. ANALOG CIRCUITS FOR VISION: THE EARLY YEARS

This idea was first explicitly raised by Horn at MIT,[10] when he proposed the use of a 2-D hexagonal grid of resistances to solve for the inverse Laplacian. This is the crucial operation in an algorithm for determining the lightness (in his case the reflectance) of objects from their image. An attempt to build an analog network for vision was undertaken by Knight in his PhD Thesis[7] for the problem of convolving images with the difference of two Gaussians, a good approximation of the Laplacian of a Gaussian filter of Marr and Hildreth.[11] The principal idea is to exploit the dynamic behavior of a resistor/capacitor transmission line, illustrated in Fig. 1. In the limit that the grid becomes infinitely fine, the behavior of the system is governed by the diffusion equation:

$$RC\frac{\partial V(x,t)}{\partial t} = \frac{\partial^2 V(x,t)}{\partial x^2}. \tag{1}$$

If the initial voltage distribution is $V(x,t=0)$ and if the boundaries are infinitely far away, the solution voltage is given by the convolution of $V(x,0)$ with a progressively broader Gaussian distribution.[7] Thus, a difference of two Gaussians can be computed by converting the incoming image into an initial voltage distribution, storing the resulting voltage distribution after a short time and subtracting it from the voltage distribution at a later time. A resistor/capacitor plane yields the same result in two dimensions. Practical difficulties prevented the successful implementation of this idea.

A different approach—exploiting CCD technology—for convolving images was successfully tried by Sage at MIT's Lincoln Laboratory,[12] based on an earlier idea of Knight.[7] In this technology, incoming light intensity is converted into a variable amount of charge trapped in potential "wells" at each pixel. By using appropriate clocking signals, the original charge can be divided by two and shifted into adjacent wells. A second step further divides and shifts the charges and so on (Fig. 2). This causes the charge in each pixel to spread out in a diffusive manner

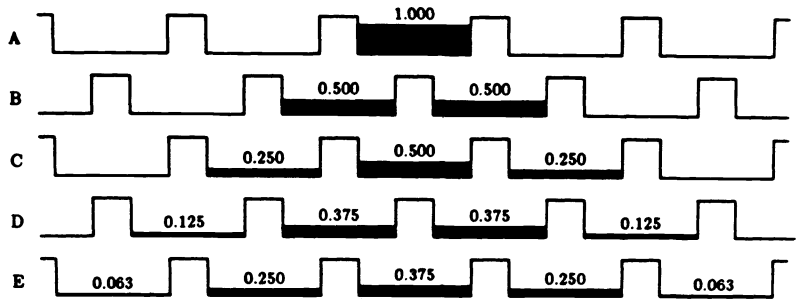


Fig. 2. Schematic of a potential "well" CCD structure evolving over time. The initial charge across the 1-D array is proportional to the incoming light intensity. The charge packet shown in **A** is then shifted into the two adjacent wells by an appropriate clocking method. Since the total charge is conserved, the charge per well is halved (**B**). In subsequent cycles, (**C, D** and **E**) the charge is further divided and shifted, resulting in a binominal charge distribution. After several steps, the charge distribution is very similar to a Gaussian distribution. One practical drawback of this scheme is that large Gaussian filters result in a very broad and flat charge distribution over many wells, making the result suceptible to device noise. From Koch.[6]

described accurately by a binominal convolution. This represents, after a few iterations, a good approximation to a Gaussian convolution. Sage extended this work to the 2-D domain[13] by first effecting the convolution in the x and then in the y direction. Their 288 by 384 pixel CCD imager convolves images (using up to 40 mixing cycles) at up to 60 times per second! Since CCD devices can be packed extremely dense—commercial CCD image sensors with one million pixel are available—such convolvers promise to be remarkably fast and area-efficient. Recently, this technique has been further extended by Yang (this volume).

## 3. ANALOG VLSI AND NEURAL SYSTEMS

The current leader in the field of analog sensory devices that include significant signal processing is undoubtedly Carver Mead at Caltech.[14] During the last 5 years he has developed a set of subcircuit types and design practices for implementing a variety of vision circuits using subthreshold analog complementary Metal-Oxide-Semiconductor (CMOS) VLSI technology. His best known design is the "Silicon" retina,[15,16] a device which computes the spatial and temporal derivative of an image projected onto its phototransistor array. The version illustrated schematically in Fig. 3a has two major components. The photoreceptor consists of a phototransistor feeding current into a circuit element with an exponential current-voltage characteristic.[17] The output voltage of the receptor $\tilde{V}$ is logarithmic over four to five orders of magnitude of incoming light intensity, thus performing automatic gain control, analogous to the cone photoreceptors of the vertebrate retina (for an up-to-date account of such phototransistors see Mann *et al.*, this volume). This voltage is then fed into a 48 by 48 element hexagonal resistive layer with uniform resistance values $R$. The photoreceptor is linked to the grid by a conductance of value $G$. An amplifier senses the voltage difference across this conductance and thereby generates an output at each pixel proportional to the difference between the receptor output and the network potential. Formally, if the voltage at pixel $i, j$ is $V_{ij}$ and the current being fed into the network at that location $I_{ij} = G(\tilde{V}_{ij} - V_{ij})$, the steady state is characterized by:

$$6V_{ij} - V_{i_1j} - V_{i+1j+1} - V_{ij+1} - V_{i-1j} - V_{ij-1} - V_{i+1j-1} = RI_{ij}. \tag{2}$$

On inspection, this turns out to be the simplest possible discrete analogue of the Laplacian differential operator $\nabla^2$. In other words, given an infinitely fine grid and the voltage distribution $V(x, y)$, this circuit computes the current $I(x, y)$ via

$$\nabla^2 V = RG(\tilde{V} - V) = RI. \tag{3}$$

The current $I$ at each grid point—proportional to $\tilde{V} - V$ and sensed by the amplifier—then corresponds to a spatially high-passed filtered version of the logarithmic compressed image intensity. Operations akin to temporal

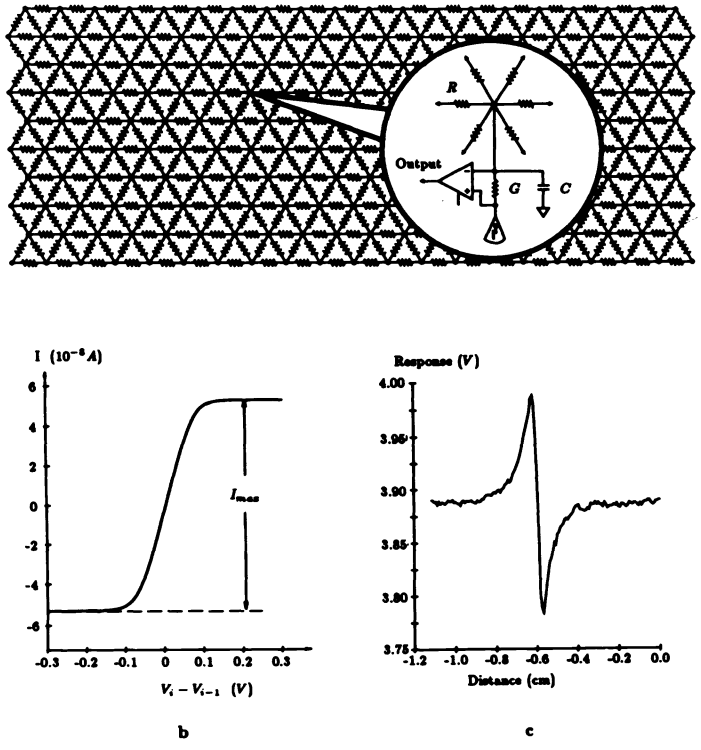


Fig. 3. The "Silicon" retina. (a) Diagram of the hexagonal resistive network with an enlarged single element. A photoreceptor, whose output voltage is proportional to the logarithm of the image intensity, is coupled—via the conductance $G$—to the resistive grid. The output of the chip is proportional to the current across the conductance $G$, or in other words, to the voltage difference between the photoreceptor and the grid. (b) The current-voltage relationship for Mead's resistive element. As long as the voltage gradient is less than $\approx 100\ mV$, the circuit acts like a linear resistive element. The output current saturates for larger voltages. The slope, and therefore the value of the resistance (determined by $I_{max}$), can be varied over 5 orders of magnitude. (c) The experimentally measured voltage response of a 48 by 48 pixel version of the retina when a step intensity edge is moved past one pixel. This response is similar to the one expected by taking the second spatial derivative of the smoothed incoming light intensity. Adapted from Mead and Mahowald.[16]

differentiation can be achieved by adding capacitive elements.[15] The required resistive elements of this circuit are designed by exploiting the current-voltage relationship (Fig. 3b) of a small transistor circuit, instead of using the resistance of a special metallic process. As long as the voltage across the device is within its linear range (a couple of 100 $mV$'s), it behaves like a constant resistance whose value can be controlled over five orders of magnitude. The current saturates for larger voltage values, a nonlinearity with very desirable effects (see below). This, then, is the basic circuit element used for all vision chips coming out of Caltech.

The response of the silicon retina to a 1-D edge projected onto the phototransistors is shown in Fig. 3c. The voltage trajectory can be well approximated by the second spatial derivative of the smoothed brightness intensity. In 2-D the response is similar to that of convolving the image with the Difference-of-a-Gaussian edge detection operator.[11] A different circuit[18] computes a simple approximation of the optical flow field induced by a spatially homogeneous motion, such as when moving a pointing device over a fixed surface (e.g., an optical mouse).

A problem plaguing analog subthreshold circuits are random offsets which vary from location to location and are caused by fluctuations in the process accuracy as well as dark currents. Such offsets, while usually not problematic for digital circuits, can be very disruptive when operating in the subthreshold analog domain, given the exponential dependency of the drain current on the gate voltage.[14] This is particularly true when computing spatial or temporal derivatives. Mead[14] (see also Mahowald, this volume) has developed a variant of the "floating gate technology" used for resetting programmable read-only-memory cells (EPROM) by means of ultra-violet light. While previously the chips were bombarded with UV radiation to erase memory, Glasser[19] demonstrated how this technology could be used to selectively write a "0" or a "1" into the cell. Mead is the first to have applied this technique to the analog domain, by building a local feedback circuit at every node of the retina (Fig. 3) which senses the local current and attempts to keep it at or near zero by charging up a capacitor located between two layers of poly positioned above each node. Exposure to UV light excites electrons suficiently to enable them to surmount the potential barrier at the silicon/silicon dioxide interface. In order to adapt the retina, a blank, homogeneous image is projected for 10-20 minutes onto the chip in the presence of UV light. This effectively creates a "floating" battery at each pixel, inducing a current exactly counteracting the effect of the offset current at that particular location. It is even possible to demonstrate after-image-like pheneomena. Mahowald (this volume) discusses this further in the case of the silicon retina.

A further potential problem in designing the type of networks discussed here is that unwanted oscillations can spontaneously arise when large populations of active elements are interconnected through a resistive grid. These oscillations can occur even when the individual elements are quite stable. Using methods from nonlinear circuit theory, Wyatt and Standley[20] have shown how this flaw can be circumvented. They have proven that if each linear active element in isolation is designed to satisfy the experimentally testable Popov criterion from control theory (which guarantees that a related operator is positive real), then stability of the overall interconnected system is guaranteed. Furthermore, their stability proof is not invalidated by the presence of any unmodelled nonlinear resistors or capacitors at unknown locations in the grid, as commonly occurs in integrated circuits.[20]

Mead's principal motivation for this work comes from his desire to understand and emulate neurobiological circuits.[14,16] He argues that the physical restrictions on the density of wires and the cost of communications imposed by the spatial layout of the electronic circuits are similar to the constraints imposed on biological circuits. Furthermore, the silicon medium provides both the computational neuroscience and engineering communities with tools to test theories under realistic, real-time conditions. To further the spread of this technology into the general academic community, all circuits are built via the silicon foundry MOSIS, even though this prevents the use of the latest circuit technology.

## 4. REGULARIZATION THEORY AND ANALOG NETWORKS

Problems in vision are usually inverse problems; the two-dimensional intensity distribution on retina or camera must be inverted to recover physical properties of the visible three-dimensional surfaces surrounding the viewer. More precisely, these problems are ill-posed in that they either admit to no solution, to infinitely many solutions, or

to a solution that does not depend continuously on the data. In general, additional constraints must be applied to arrive at a stable and unique solution. One common technique to achieve this, termed "standard regularization",[21] is via minimization of a given "cost" functional (for earlier examples of this see[22–25]). The first term in these functionals assesses by how much the solution diverges from the measured data. The second term measures how closely the solution conforms to certain *a priori* expectations, for instance that the final surface should be as smooth as possible. Let us briefly consider the problem of fitting a 2-D surface through a set of noisy and sparse depth measurements, a well-explored problem in computer vision.[22] Specifically, a set of sparse depth measurements is given on a 2-D lattice, $d_{ij}$, which are assumed to be corrupted by some noise process. It is obvious that infinitely many surfaces, $f_{ij}$, can be fitted through the sparse data set. One way to regularize this problem is to find the surface $f$ that minimizes

$$\sum_{i,j}((f_{i+1j} - f_{ij})^2 + (f_{ij+1} - f_{ij})^2) + \lambda \sum_{i,j}(f_{ij} - d_{ij})^2, \tag{4}$$

in which $\lambda$ depends on the signal-to-noise ratio and the second sum only contains contributions from those locations $i$ where data exist. Equation (4) represents the simplest possible functional, even though many alternatives exist.[22,24,26] This and all other quadratic regularized variational functionals of early vision can be solved within simple linear resistive networks by virtue of the fact that the electrical power dissipated in linear networks is quadratic in the current or voltage.[21,27] The resistive network will then converge to its unique equilibrium state in which the dissipated power is at a minimum (subject to the source constraint). The static version of this statement is known as *Maxwell's Minimum Heat Theorem*. For instance, the steady-state of the resistive network in Fig. 4a minimizes eq. (4) if the voltage $V_{ij}$ is identified with the discretized solution surface $f_{ij}$, the battery $E_{ij}$ with the data $d_{ij}$, and the ratio of the variable conductance-to-ground $G_{ij}$ to the constant horizontal resistance $R$ with $\alpha$ (for more details, see Fig. 4). The power minimized by this circuit is then formally equivalent to the functional of eq. (4). The performance of an experimental 20 by 20 subthreshold, analog CMOS VLSI circuit is illustrated in Fig. 4.[28] For an infinitely fine grid and a voltage source $E(x, y)$ the surface interpolation chip computes the voltage distribution $V(x, y)$ according to the modified Poisson equation (see also[9]):

$$\nabla^2 V + RGV = RGE, \tag{5}$$

with either an arbitrary Dirichlet boundary conditions (such as zero voltage along the boundary) or a zero voltage slope (that is, no current across the boundary) Neumann boundary condition. If $RG$ is a constant across the grid, this equation is sometimes known as the Helmholtz equation. Note that the difference to eq. (3) lies in the choice of observable, current ($I = \tilde{V} - V$) versus voltage ($V$).

A large number of problems in early vision, such as detecting edges, computing optical flow, estimating depth from two images etc. have a similar architecture, with resistive connections among neighboring nodes implementing the constraint that objects in the real world tend to be smooth and continuous. For instance, an algorithm for estimating depth from two images—exploiting an analogy between binocular stereo and optical flow—can be solved within a linear resitive network nearly identical to the one shown in Fig. 4a.[29] This method uses the gradient of the image brightness to directly compute depth. It should be emphasized, however, that in some cases, such as computing the optical flow, the resistances in the network depend on data and may even be negative, making hardware implementation a nontrivial effort.

## 5. DISCONTINUITIES

The most interesting locations in any scene are arguably those locations at which some feature changes abruptly, for instance the 2-D optical flow at the boundary between a moving figure and the stationary background or the color across the sharp boundaries in a painting. Geman and Geman[30] (see also[31]) introduced the powerful concept of a binary line process at location $i, j$, $l_{i,j}$, which explicitly codes for the absence ($l_{ij} = 0$) or presence ($l_{ij} = 1$) of a discontinuity. Further constraints, such that discontinuities should occur along continous contours (as they do, in general, in the real world) or that they rarely intersect, can be incorporated into their theory, which is based on a statistical estimation technique (see also[32]). In the case of surface interpolation and smoothing, maximizing the

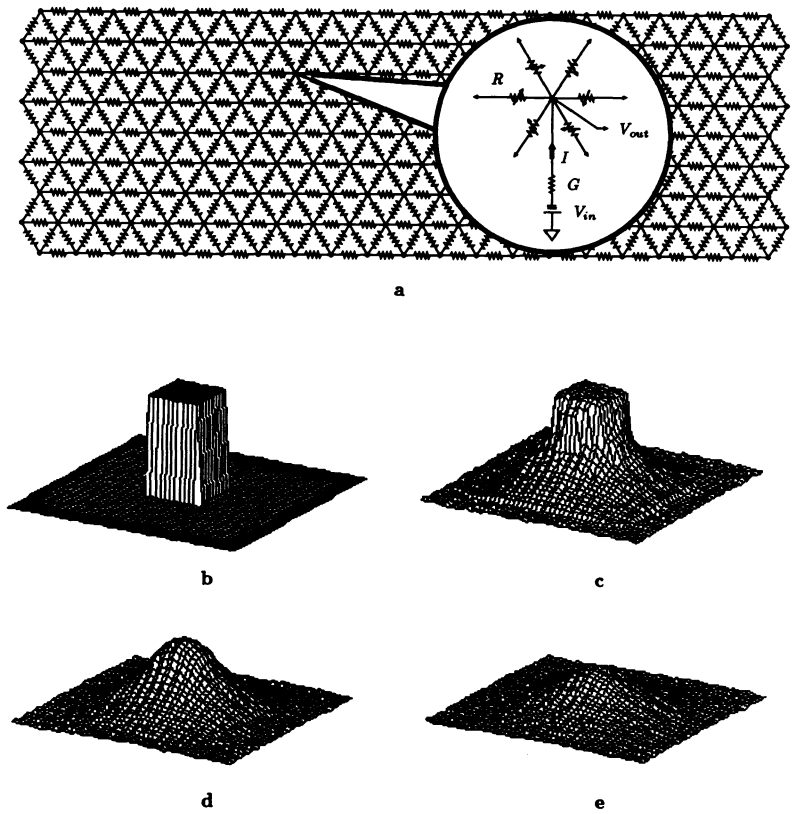


Fig. 4. Surface interpolating network. **(a)** At those locations where depth data are avaliable, the values of the battery $V_{in}$ and of the conductance $G$ are set to their appropriate values, via additional sample-and-hold circuitry. The horizontal resistances $R$ are built using the nonlinear circuit of Fig. 3b. The output is the voltage $V_{out}$ at each location, corresponding to the reconstructed "smooth" surface. This circuit solves, for small enough voltage gradients, Poisson's equation with a zero-slope boundary condition, via minimization of eq. (4). Experimental results from a 48 by 48 pixel analog CMOS VLSI circuit are shown next.[28] **(b)** The input voltage $V_{in}$, corresponding to a flat, 2-pixel-wide strip around the periphery and a central 4-pixel-wide tower (solid coloring). At these locations, the conductance $G$ is set to a constant, fixed value, while $G$ is zero everywhere else. Thus, no data are present in the area between the bottom of the tower and the outside strip. **(c)**, **(d)** and **(e)** show the output voltage for a high, medium, and low value of the transversal resistance $R$. If $R$ is small enough, the resulting smoothing will flatten out the central tower. From Koch.[6]

*a posteriori* estimate of the solution can be shown to be equivalent to minimizing:

$$\sum_{i,j}(1-l_{ij}^h)(f_{i+1j}-f_{ij})^2+\sum_{i,j}(1-l_{ij}^v)(f_{ij+1}-f_{ij})^2+\lambda\sum_{i,j}(f_{ij}-d_{ij})^2+\alpha\sum_{i,j}V(l_{ij}), \tag{6}$$

where $l_{ij}^h$ and $l_{ij}^v$ are the horizontal and vertical depth discontinuities, $\lambda$ depends on the signal-to-noise ratio, $\alpha$ is a fixed parameter, and $V$ a potential function containing a number of terms penalizing or encouraging specific configurations of line processes (for instance that they should occur along continuous contours). In the case of 1-D surface interpolation, a simple example is $V(l_i)=l_i$. In other words, the line process $l_i$ between $i$ and $i+1$ will be set to 1 if the "cost" for smoothing, that is $(f_{i+1}-f_i)^2$, is larger than the parameter $\alpha$. Otherwise, $l_i=0$. In 2-D, this potential function can be considerably more complex.[32,33] Discontinuities greatly improve the performance of early vision processes, since they allow algorithms to smooth over unreliable or sparse data as well as account for boundaries between figures and ground. Within the last years, they have been used to demarcate boundaries in the intensity, color, depth, motion, and texture domains.[3,30–35] Figure 5 demonstrates their advantages in the case of computing the optical flow field induced by moving objects.

Line discontinuities can be implemented in at least two different ways. In a *hybrid* implementation, each line process is represented by a simple binary switch.[36] When the switch (representing $l_{ij}^h$) is open, no current flows across the connection between the two adjacent nodes $i,j$ and $i+1,j$, no matter what the voltage gradient between the two locations. If the switch is closed, i.e. $l_{ij}^h=1$, the normal interaction between the values at points $i,j$ and $i+1,j$ occurs. The network operates by switching between two distinct modes. In the analog cycle the network settles into the state of least power dissipation, *given a fixed distribution of switches.* In the digital phase, the line processes are evaluated using eq. (6); i.e., the switches are set to the state minimizing eq. (6).

A much more elegant and completely analog implementation combines the binary line process with a linear resistance within a single two-terminal nonlinear circuit device.[35,37] The key idea behind this "resistive fuse" is simple. Let us assume that we are discussing a particular physical process, e.g., surfaces. If the voltage gradient across the device—corresponding to the difference in depth between two neighboring values of the surface—is small, e.g., less than $\alpha^{1/2}$ to stay within the previous example of eq.(6), then the algorithm assumes that both points lie on the same surface but are corrupted by noise. In that case the device will act as a resistance and will conduct a current proportional to the voltage gradient. If however, the voltage across the device is above a threshold (e.g., $\alpha^{1/2}$), the algorithm assumes that the two points lie on two different surfaces and no smoothing occurs, that is, the device open-circuits and no current flows. Figure 6 shows the measured I-V relationship for an analog version of such a resistive fuse.[37] Since usually no *a priori* information exists as to the optimal value of $\alpha$, it is best to adopt a deterministic annealing method for varying $\alpha$, closely related to continuation methods for minimizing the associated non-convex variational functional of eq. (6).[38] The notion of minimizing power in linear networks implementing quadratic regularization algorithms must be replaced by the more general notion of minimizing the total co-content $J$ for these networks, where $J=\int_0^V f(V')dV'$ for a resistor defined by $I=f(V)$.[39] Figure 7 illustrates the performance of an experimental test circuit, demonstrating its robust performance.

The flow field shown in Fig. 5 was evaluated by having a conventional digital computer simulate the appropriate resistive network with fuses, in this case, two networks of the type shown in Fig. 4 and corresponding to the $x$ and $y$ components of the optical flow field, interconnected with linear resistances. The flow field—induced by the time-varying image intensity $I(x,y,t)$—is regularized using a first-order smoothness constraint.[23] The amount of smoothing is governed by the constant conductance value $R$ of the upper and lower horizontal grids.

## 6. ANALOG CHIPS VERSUS DIGITAL COMPUTERS

As we have seen, all of the above circuits exploit the physics of the system to perform operations useful from a computational point of view. Thus, the transient voltage or charge distribution at some time in the networks of Figs. 1, 2 or 3 corresponds to the solution, in this case convolution of the image intensity with a Difference-of-a-Gaussian. In the networks derived from standard or non-standard regularization theory, the stationary voltage distribution corresponds to the interpolated surface (Fig. 4) or to the optical flow (Fig. 5). These quantities are gov-

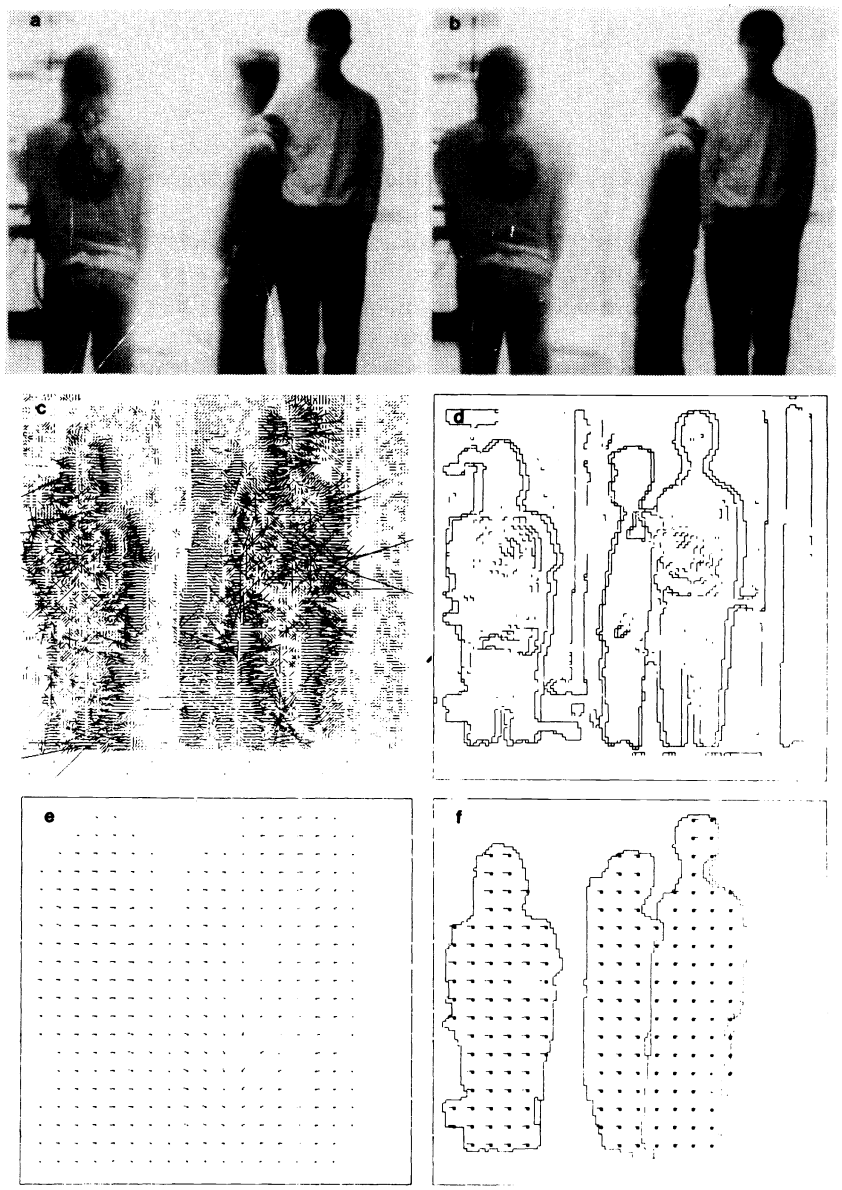


Fig. 5. Optical flow with motion discontinuities. **(a)** and **(b)** show two 128 by 128 pixel images captured by a video camera. The two leftward people move toward the left, while the rightward person moves to the right. **(c)** The initial local velocity data prior to smoothing. **(d)** The zero-crossings of the Laplacian of a Gaussian of both images. The zero-crossings are thresholded to remove noise. **(e)** Smooth and subsampled optical flow obtained by solving the associated variational functional.[23] Two undifferentiated "blobs" move to the left and one moves to the right. **(f)** The subsampled optical flow computed with an analog, continuous form of the "resistive fuse". In order to visualize the behavior of the fuses, we indicate their state with a solid line if the voltage difference across them exceeds some threshold value. The fuses are prevented from "breaking" at locations where no zero-crossings are present. The final optical flow clearly indicates three moving people and was obtained by solving for the steady-state of the associated non-linear resistive network. From Harris.[35]

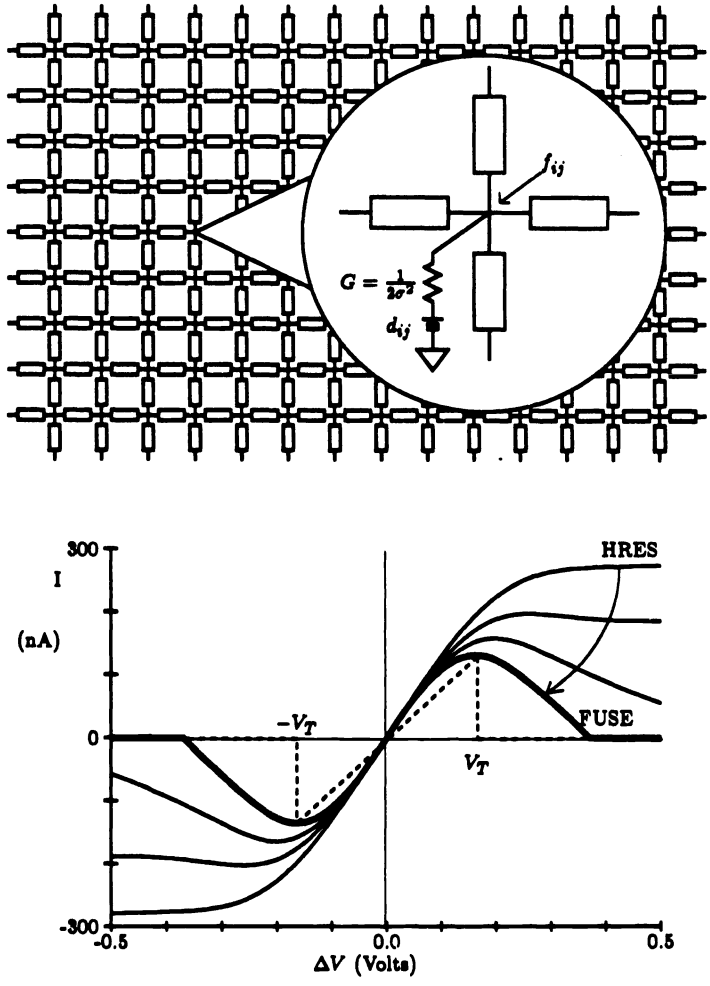


Fig. 6. **(a)** Schematic diagram of a 20 by 20 pixel surface interpolation and smoothing chip. A rectangular mesh of resisitive fuse elements (shown as rectangles) provides the smoothing and segmentation ability of the network. The data are given as battery values $d_{ij}$ (eq. 6) and the conductance $G$ depends on the variance $\sigma^2$ of the additive Gaussian noise assumed to corrupt the data. If no data are available, $G = 0$. The output is the voltage $f_{ij}$ at each node (compare Fig. 4a). Parasitic capacitances (not shown) provide the dynamics. The steady-state of the circuit corresponds to one of the local minima of the non-convex variational functional of eq. (6). **(b)** Measured $I - V$ relation for different settings of the resistive fuse. The $I - V$ curve can be continuously varied from the hyperbolic tangent of Mead's saturating resistor (HRES) to that of an analog fuse. The $I - V$ curve of a binary fuse is also indicated (dashed line). For a voltage of less than $V_T = \sqrt{\alpha}$ across this two-terminal device, the circuit acts as a resistor with fixed conductance. Above $V_T$, the current is either abruptly set to zero (binary fuse) or smoothly goes to zero (analog fuse). Independent voltage control lines allow real-time changes of both the slope (over four orders of magnitude) as well as $V_T$ (over one order of magnitude). From Harris.[37]

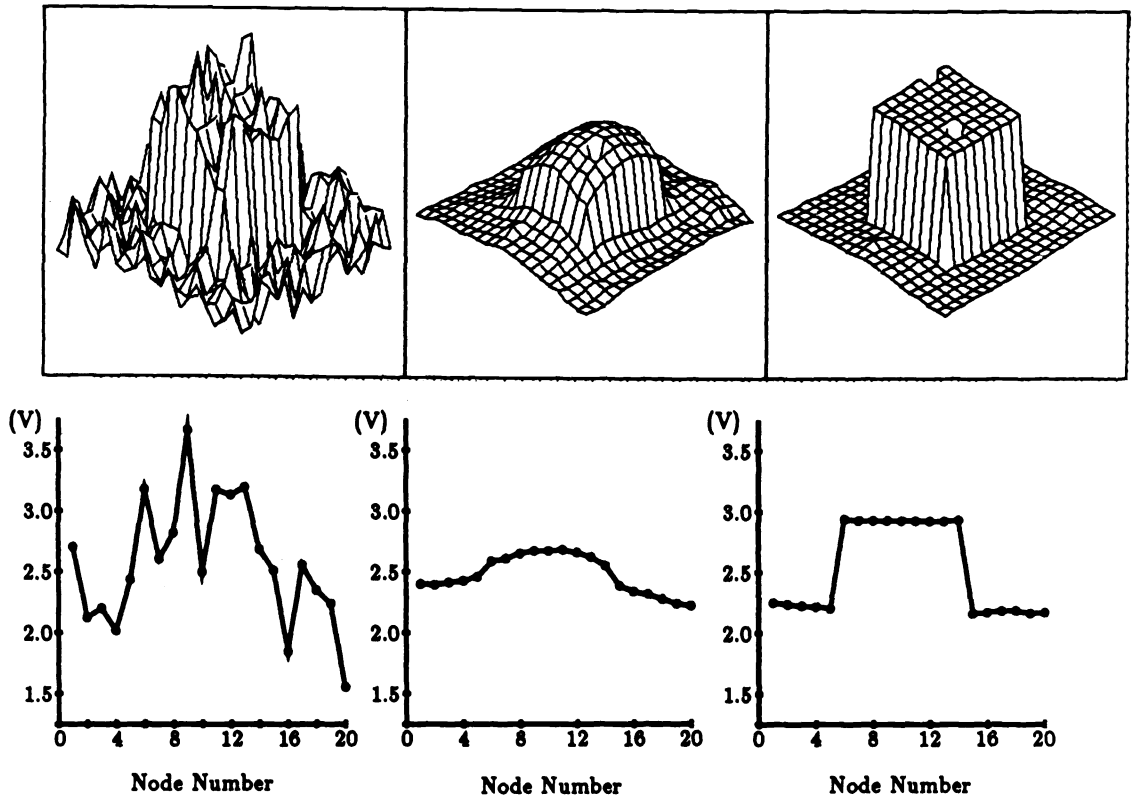


Fig. 7. Experimental data from the chip described in Fig. 6. Similar to Fig. 4, a tower is used as input data, with the height of the tower set to 3 $V$ rising from a plane of 2 $V$ with superimposed Gaussian noise. Note however, that different from the case illustrated in Fig. 4, data exist everywhere. **(a)** The input with the variance of the noise set to 0.2 $V$. **(b)** The voltage output with the fuse configured as a saturating resistor. **(c)** The voltage output when the $I-V$ curve of the fuse has been changed from the saturating resistor to that of the analog fuse (following the arrow in Fig. 6B) while the conductance has been increased. **(d-f)** The same behavior along a horizontal slice across the chip for an increased level of noise ($\sigma^2 = 0.4\ V$, corresponding to 40% of the height of the central tower). The smoothing and segmentation abilities of the fuse are obvious. Cooperativity and hysteresis are required for this optimal performance. Notice the "'bad" pixel in the middle of the tower (in **c**). Its effect is localized, however, to a single element. From Harris.[37]

erned by *Kirchhoff's* laws, instead of being symbolically computed via execution of software in a digital computer. Furthermore, the architecture of the analog resistive circuits reflects the nature of the underlying computational task, for instance, smoothing, while digital computers—being Turing universal—do not. This difference has interesting implications for structure-function relationships in neurobiology. One of the advantages of analog circuits is that their operating mode is optimally suited to analog sensory data since they avoid any temporal aliasing problems. Furthermore, their robustness to imprecisions or errors in the hardware, their processing speed and low power consumption (Mead's retina requires less than 1 $mW$, most of which is used in the photo-conversion stage) and their small size make analog smart sensors very attractive for tele-robotic applications, remote exploration of planetary surfaces, and a host of industrial applications where their power hungry, heat producing, bulky and slow digital cousins are unable to compete.

The two principal drawbacks of analog VLSI circuits are their lack of flexibility and their imprecision. The above circuits are all hardwired to perform very specific tasks, unlike digital computers which can be programmed to approximate any logical or numerical operation. Only parameters associated with this algorithm, for instance the smoothness in the case of Figs. 4 and 5, can be varied. Thus, digital computers appear vastly preferable for developing and evaluating new algorithms; analog implementations should only be attempted after such initial exploration of algorithms. Furthermore, although 12- and even 16-bit analog-to-digital converters are commercially available, it seems unlikely that the precision of analog vision circuits will exceed 7 to 8 bits in the next years. However, for a number of important tasks, such as navigation or tracking, the incoming intensity data are rarely more accurate than 1%. Moreover, it seems unlikely that the individual circuit components in biological systems, neurons, process data with more than, at most, 100 levels of resolution, that is between 6 and 7 bits.

## 7. THE FUTURE

Within the last year, a number of exciting developments have occured which bode well for the future of analog vision circuits. Mahowald and Delbrück[40] from Mead's laboratory have built and tested an analog CMOS VLSI circuit implementing a version of Marr and Poggio's[41] cooperative stereo algorithm. Two 1-D phototransistor arrays, with 40 elements each, located next to each other on the chip provide the input to the circuit. A winner-take-all circuit (as proposed by Prazdny[42]) selects the most active node among the nine possible disparity values at each pixel, replacing the inhibitory interaction in the original algorithm. In this volume, Hakkarainen, Little and Wyatt report on the CCD version of a related stereo algorithm.

One problem with the resistive networks derived from regularization theory is that the values of the individual circuit elements, such as conductances or voltage sources, can depend on the measured data or even be negavtive in value, raising problems with network stability. In the case of computing the optical flow (fig. 5), the value of the conductance linking the top and bottom grids is given by the product of the spatial derivative $G = -I_x I_y$. Harris, formerly at MIT and now at Caltech, has shown how this problem can be circumvented via the use of so-called "constraint boxes", which impose a generalized constraint equation.[26,43] For the case of reconstructing surfaces using a smoother functional than the one of eq. (4) (the biharmonic equation instead of Laplace's or Poisson's equation), this three-node reciprocal circuit implements an equation of the form $V_a - V_b - V_c = 0$. The device is unusual in that all of its terminals can act as input or output nodes. Thus, if nodes $a$ and $b$ are held constant, then the $c$ node is fixed to $V_a - V_b$. Using these constraint boxes in the case of computing smooth optical flow,[23,33] all resistance values are positive and data-independent, a considerable advantage when building these circuits. In Harris's subthreshold analog CMOS circuit implementation (via MOSIS) of these constraint networks, their principal drawback is their large size, currently precluding their incorporation into 2-D networks.

A team at MIT headed by J. Wyatt, and including B. Horn, H.-S. Lee, T. Poggio and C. Sodini, is initiating an ambitious effort to fabricate analog, early vision chips exploiting different circuit technologies, such as CCD or mixed bipolar and CMOS devices. They plan to build various 2-D spatial correlator and convolver circuits, analog image memories as well as motion sensors. As part of this effort, new methods for estimating first and second image moments, that is for computing both the center of "gravity" as well as the axis of least inertia have been developed and successfully placed onto CMOS chips[44] (Standley, Wyatt and Horn, this volume).

Fusion of information is being attempted in Koch's laboratory at Caltech, by integrating a set of simple resistive networks computing depth and depth discontinuities, with circuits locating edges and optical flow. Initially, only 1-D designs of these networks are attempted, greatly reducing the technical difficulties involved. A chip with 64 photoreceptors computing thresholded zero-crossings using the difference between two 1-D resistive grids has been built and works in a very robust manner[45] (and this volume). Moreover, a number of different circuits estimating the 1-D optical flow using both the gradient as well as the correlation methods have been designed and tested[46] (e.g., Bair and Koch, this volume; Moore and Koch, this volume). Furthermore, in order to demonstrate that such analog chips with on-board photoreceptor arrays are robust enough to work in a real-time environment, a number of vision chips have been placed onto small mobile vehicles navigating auntonomously in a laboratory environment.[47]

A number of other laboratories, such as Rockwell's International Science Center, are also engaged in efforts to build early vision sensors (Mathur, Liu, Wang, Koch, and Luo, this volume). Thus, it appears that the analog computers of the 1940s and 1950s (for a discussion of their use see[48]), until recently considered extinct, are making a limited sort of comeback in the form of highly dedicated smart, fast and power-lean vision chips.

## 8. ACKNOWLEDGMENTS

Research on early vision chips in our laboratory is supported by the the National Science Foundation, the Office of Naval Research, Rockwell International Science Center and the Artificial Intelligence Center at Hughes Aircraft.

## 9. REFERENCES

1. B. K. P. Horn, *Robot Vision*, MIT Press, Cambridge, 1986.

2. H. G. Barrow and J. M. Tenenbaum, "Computational vision," *Proc. IEEE* **69:** 572–595, 1981.

3. T. Poggio, E. B. Gamble, and J. J. Little, "Parallel integration of vision modules," *Science* **242:** 436–440, 1988.

4. W. D. Hillis, *The Connection Machine*, MIT Press, Cambridge, 1985.

5. J. Little, "Parallel algorithms for computer vision on the connection machine," *Artif. Intell. Lab Memo* **No. 928**, MIT, Cambridge, 1987.

6. C. Koch, "Resistive networks from computer vision: A tutorial," in: *An Introduction to Neural and Electronic Networks*, S. F. Zornetzer, J. C. Davis, and C. Lau (eds.), pp. 293–305, Academic Press, New York, 1990.

7. T. Knight, *Design of an Integrated Optical Sensor with On-Chip Preprocessing*, PhD Thesis, Dept. of Electrical Engineering and Computer Science, MIT, Cambridge, 1983.

8. N. Strausfeld, *Atlas of an Insect Brain*, Springer, Heidelberg, 1975.

9. G. R. Nudd, P. A. Nygaard, G. D. Thurmond, and S. D. Fouse, "A charge-coupled device image processor for smart sensor applications," *SPIE Image Understanding Systems & Industrial Applications* **155:** 15–22, 1978.

10. B. K. P. Horn, "Determining lightness from an image," *Compt. Graph. Imag. Processing* **3:** 277–299, 1974.

11. D. Marr and E. C. Hildreth, "Theory of edge detection," *Proc. Roy. Soc. Lond. B* **207:** 187–217, 1980.

12. J. P. Sage, "Gaussian convolution of images stored in a charge- coupled device," *Quart. Tech. Rep.*, August–October 1983, MIT Lincoln Laboratory, Lexington, 1984.

13. J. P. Sage and A. L. Lattes, "A high-speed two-dimensional CCD Gaussian image convolver," *Quart. Tech. Rep.*, August–October 1986, MIT Lincoln Laboratory, Lexington, 1987.

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

15. M. A. Sivilotti, M. A. Mahowald, and C. A. Mead, "Real-time visual computation using analog CMOS processing arrays," in: *1987 Stanford Conf. VLSI*, pp. 295–312, MIT Press, Cambridge, 1987.

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

17. C. A. Mead, "A sensitive electronic photoreceptor," in: *1985 Chapel Hill Conf. on Very Large Scale Integration*, pp. 463–471, 1985.

18. J. E. Tanner, *Integrated Optical Motion Detection*, PhD Thesis, Dept. of Computer Science, 5223:TR:86, Caltech, 1986.

19. L. A. Glasser, "A UV write-enable PROM," in: *1985 Chapel Hill Conf. on VLSI*, W. Fuchs (ed.), pp. 61–65, Computer Science Press, Rockville, MD, 1985.

20. J. L. Wyatt and D. L. Standley, "Criteria for robust stability in a class of lateral inhibition networks coupled through resistive grids," *Neural Computation* **1:** 58–67, 1989.

21. T. Poggio, V. Torre, and C. Koch, "Computational vision and regularization theory," *Nature* **317:** 314–319, 1985.

22. W. E. L. Grimson, *From Images to Surfaces*, MIT Press, Cambridge, 1981.

23. B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artif. Intell.* **17:** 185-203, 1981.

24. D. Terzopoulos, "Multilevel computational processes for visual surface reconstruction," *Comp. Vision Graph. Image Proc.* **24:** 52–96, 1983.

25. E. C. Hildreth, *The Measurement of Visual Motion*, MIT Press, Cambridge, 1984.

26. J. G. Harris, "A new approach to surface reconstruction: the coupled depth/slope model," *Proc. IEEE 1. Intl. Conf. Comp. Vision*, pp. 277–283, London, 1987.

27. T. Poggio and C. Koch, "Ill-posed problems in early vision: from computational theory to analogue networks," *Proc. R. Soc. Lond. B* **226:** 303–323, 1985.

28. J. Luo, C. Koch, and C. Mead, "An experimental subthreshold, analog CMOS two-dimensional surface interpolation circuit," *Neural Information and Processing Systems Conf.*, Denver, November 1988.

29. A. K. Chhabra and T. A. Grogan, "Estimating depth from stereo: variational methods and network implementation," in: *IEEE Intl. Conf. Neural Networks*, San Diego,1988.

30. S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images," *IEEE PAMI* **6:** 721–741, 1984.

31. A. Blake and A. Zisserman, *Visual Reconstruction*, MIT Press, Cambridge,1987.

32. J. Marroquin, S. Mitter and T. Poggio, "Probabilistic solution of ill- posed problems in computational vision," *J. Am. Statistic Assoc.* **82:** 76–89, 1987.

33. J. Hutchinson, C. Koch, J. Luo, and C. Mead, "Computing motion using analog and binary resistive networks," *IEEE Computers* **21:** 52–63, 1988.

34. D. Terzopoulos, *IEEE Trans. Pattern Anal. Machine Intell.* **8:** 413–424, 1986.

35. J. G. Harris, C. Koch, E. Staats, and J. Luo, "Analog hardware for detecting discontinuities in early vision," *J. Intl. Comp. Vision* **4:** 211–223, 1990.

36. C. Koch, J. Marroquin, and A. Yuille, "Analog 'neuronal' networks in early vision," *Proc. Natl. Acad. Sci. USA* **83:** 4263–4267, 1986.

37. J. G. Harris, C. Koch, and J. Luo, "A two-dimensional analog VLSI circuit for detecting discontinuities in early vision," *Science* **248:** 1209–1211, 1990.

38. J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.

39. J. G. Harris, C. Koch, J. Luo, and J. Wyatt, "Resistive fuses: Analog hardware for detecting discontinuities in early vision," in: *Analog VLSI Implementations of Neural Systems*, C. Mead and M. Ismail (eds.), pp. 27–56, Kluwer, Norwell, MA, 1989.

40. M. Mahowald and T. Delbrück, "An analog VLSI implementation of the Marr-Poggio stereo correspondence algorithm," *Proc. Intl. Neural Network Soc.*, Boston, 1988.

41. D. Marr and T. Poggio, "Cooperative computation of stereo disparity," *Science* **194:** 283–287, 1976.

42. K. Prazdny, "Detection of binocular disparities," *Biol. Cybern.* **52:** 93–99, 1985.

43. J. G. Harris, "An analog VLSI chip for cubic spline surface interpolation," *Neural Information and Processing Systems Conf.*, Denver, November 1988.

44. B. K. P. Horn, "Parallel networks for machine vision," *Artif. Intell. Lab. Memo* **No. 1071**, MIT, Cambridge, 1989.

45. W. Bair and C. Koch, "An analog VLSI chip for finding edges from zero-crossings," in: *Neural Information Processing Systems*, D. Touretzky (ed.), Morgan Kaufmann, Palo Alto, CA, in press 1991.

46. T. Horiuchi, J. Lazzaro, A. Moore, and C. Koch, "A delay-line based motion detection chip," in: *Neural Information Processing Systems*, D. Touretzky (ed.), Morgan Kaufmann, Palo Alto, CA, in press 1991.

47. C. Koch, W. Bair, J. G. Harris, T. Horiuchi, A. Hsu, and J. Luo, "Real- time computer vision and robotics using analog VLSI circuits," in: *Advances in Neural Information Processing Systems, Vol. 2*, D. Touretzky (ed.), pp. 750–757, Morgan Kaufmann, Palo Alto, CA, 1990.
48. W. J. Karplus, *Analog Simulation: Solution of Field Problems*, McGraw-Hill, New York, 1958.