An Active Resistor Network for Gaussian Filtering of Images

Haruo Kobayashi, Joseph L. White, *Student Member, IEEE,* and Asad A. Abidi, *Member, IEEE*

Abstract —The architecture of an active resistive mesh containing both positive and negative resistors to implement a Gaussian convolution in two dimensions is described. With an embedded array of photoreceptors, this may be used for image detection and smoothing. The convolution width is continuously variable by 2:1 under user control. Analog circuits implement a 45×40 mesh on a 2-μm CMOS IC, and perform an entire convolution in 20 μs on applied images.

I. Introduction

HARDWARE capable of sensing an input in two dimensions and processing it in parallel to obtain results in real time is of great interest in applications such as low-power compact image recognition systems. In digital signal processors today, a 2D input from a sensor is first scanned and quantized, and subsequently processed using pipelined parallel algorithms to obtain a fast throughput rate [1]. The data at each grid point in the 2D input, corresponding to one pixel in the case of a sampled image, serially enter this signal processor and flow through it at some usually fast clock rate. A substantial increase in throughput may be obtained over this signal flow rate by using *simultaneous* processing *per pixel*, particularly if the signal fan-out is eliminated by not digitizing the input but retaining it as an analog quantity. This is how signal processing takes place in natural biological systems [2]–[4].

Much of signal processing consists of data reduction and the extraction of high-level content for purposes such as identification, classification, or storage. The hardware to accomplish this will very often implement an algorithm derived from a study of physical or biological systems, which naturally perform a similar task. In a programmable digital signal processor, an explicit algorithm is entered as a sequence of instructions, or as their hardwired equivalent in a dedicated processor. Analog hardware, on the other hand, cannot be programmed as

digital operations may be, and is almost always hardwired: a circuit must be constructed in which Kirchhoff's laws and the terminal characteristics of the components together embody the desired algorithm. Insofar as this synthesis is guided by experience, ingenuity, and taste, the approach is ad hoc and limited in its generality; but when successfully executed, it may offer a savings in power and enhancement in speed by orders of magnitude over the digital approach [5]. The input to an analog signal processor is some current or voltage, the output some other voltage or current determined by the laws of physics governing the circuit. The early analog computers were built on this principle, but being composed of building blocks with quite general functions, they were not very efficient in hardware for massively parallel tasks. Translinear integrated circuits are one well-known example of an efficient use of hardware to embody complex nonlinear algorithms, although usually for scalar or one-dimensional array inputs. They achieve hardware efficiency by exploiting transistor device physics rather than from complex building blocks such as operational amplifiers; they are also hardwired to accomplish a specific task [6], [7]. Our work deals with a class of circuits suited to simultaneous signal processing in two dimensions also using processing at the transistor level.

II. Image Smoothing using Simultaneous 2D Signal Processing

This section will discuss the algorithm and architecture of a particular image processing function we have implemented for potential use in compact machine vision systems [8].

A. Smoothing Images by a Gaussian Operation

Many electronic image recognition systems tend to replicate the hierarchy from low- to high-level processing found in biological organisms. A raw image is usually smoothed to suppress noisy features; its outline is then obtained with some form of edge-enhancement operation, and the outline after normalization and rotation is compared with stored templates. While the quantity of data might reduce along this chain, the complexity of the operations increases significantly. Our work relates to the lowest level of image processing, the smoothing of raw

image data with a Gaussian convolution function of variable width.

There is broad evidence suggesting that a noisy image is best smoothed by a Gaussian convolution kernel prior to edge enhancement. This corresponds to the defocusing action of a lens, and is inherent in many biological systems. The defocusing blurs the small sharp features characteristic of visual noise, which are extraneous to important objects in the field of view. Unless the image is properly smoothed beforehand, differentiating the intensity map of the image to enhance the edges will also accentuate the sharp noisy features. Theoretical work has proven that a noisy image is best smoothed by a Gaussian convolution kernel to obtain the largest signal-to-noise ratio after differentiation [9], [10].

The optimal width, or extent, of the convolution used to smooth a particular image depends on the spatial standard deviation of the noise, and also on the scale of the objects which is usually not known in advance. The width of the Gaussian smoothing must therefore be variable under the control of the user. Adaptive methods such as scale space filtering [11] rely on this capability. Our experiments suggest that a Gaussian with a width variable by a factor of 2 is adequate to smooth the noise in many simple images sampled at a resolution of 50 by 50 pixels.

We set about after these considerations to implement one analog integrated circuit capable of sampling an image at a resolution of 50 pixels on a side, smoothing it by a Gaussian in about 5 μs, and giving the user the flexibility of continuously varying the Gaussian width by a factor of 2:1. This speed of operation is orders of magnitude faster than digital implementations of this convolution function, which in addition to the requirements of image buffering also require the image to be circulated several times through a filter to obtain the property of variable width.

B. Computation in 2D Using Resistive Meshes

Resistor networks were used as analog computers in the past to solve complex boundary value problems in electromagnetics [12]–[15]. These were later replaced by numerical simulation on digital computers, primarily because of the ease of programmability. Digital computation, however, could neither surpass the low power dissipation nor the speed of analog computers, because when the latter solve complex 2D problems, the currents and voltages could attain their final values within a very short RC relaxation time. This high speed is the main attraction of analog computation for 2D real-time signal processing, in that the number of calculations unlike digital computation does not grow proportionally to the resolution, but more as the square root. The use of this concept for similar applications has also been noted elsewhere [16].

Unlike a resistive sheet subject to a potential difference between two edges, where the resulting lateral equipotential contours solve electrostatic or magnetostatic field
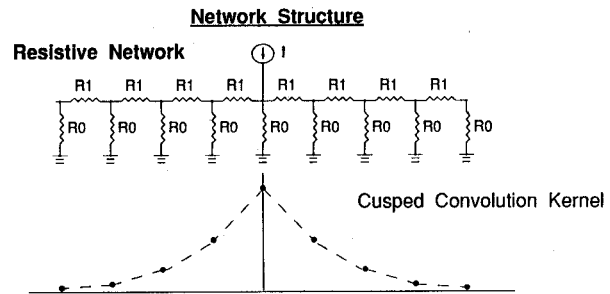

Fig. 1. 1D mesh with leakage resistors to ground, and its convolution kernel.

problems, the contours in a sheet which also has a continuous *leakage* to ground will decay in a characteristic fashion in response to a voltage applied at a single point. The spatial rate of decay depends on the leakage conductivity to ground relative to the lateral conductivity. This decay function may be thought of as the spatial impulse response of the leaky resistive sheet, or, equivalently, its convolution kernel; the potential contours in response to multiple-point stimuli will then be determined by linear superposition. Consider, for example, a one-dimensional discrete version of the leaky resistive sheet composed of a uniform linear mesh of resistors R_1 with resistors R_0 from every node to ground (Fig. 1). In response to a current excitation at one node, the resulting voltage distribution on the mesh decays n nodes away from the excitation according to an exponential function $\exp(-nR_1/R_0)$ [16]. This convolution kernel differs from a Gaussian in two important ways: it has a slower decay at its tails, and the exponentials on either side of the excitation meet at the center to produce a cusp (Fig. 1). The discontinuity in derivative at this point would produce undesirable results when this function is applied to a noisy image and then followed by edge enhancement. The mesh must therefore be modified to produce a characteristic function which better resembles the flat-topped Gaussian at the point of excitation. Obtaining a practical realization of this mesh was one of the key contributions of our work.

C. An Active Resistive Mesh Implementing Gaussian Convolution

We first qualitatively examine why the resistive mesh in the previous example produces a cusped convolution kernel, and how it must be modified. An indirect procedure for synthesizing the desired network is then described, followed by methods to extend it to two dimensions.

The spatial derivative of voltage at a point in a resistive sheet or discrete mesh specifies the potential gradient or the electric field there. According to the point form of Ohm's law, $J = \sigma E$, a current injected at a point (assuming the point has nonzero extent, so that the current density there is not infinite) on a resistive sheet with leakage to ground will produce some nonzero electric field (E) there, and therefore a nonzero potential gradient. A nonzero J may produce a zero E only if $\sigma \rightarrow \infty$,

which implies that the sheet must appear perfectly conductive at the point of injection. If a negative resistance is introduced to locally neutralize the dissipation in the sheet, while maintaining the dissipation across the large scale, a convolution function may be obtained with a flat top and decaying tails. It is plausible to achieve this in a discrete resistive mesh by introducing negative resistors not between every node, because that would simply modify the value of R_1, but between every other node, or perhaps even straddling several nodes. Investigating this numerically, we found that a mesh implementing a convolution of the desired shape could be obtained using negative resistors of a certain value connecting nodes with their *second nearest* neighbors. We also came upon an alternative procedure to synthesizing the same mesh, based on the theoretical work relating to the optimal smoothing of images. This is now described.

Poggio *et al.* [9] have analyzed how to smooth samples V_j, $-\infty < j < \infty$, of a noisy function to best estimate the derivative if the noise were not present. They seek a fitting function $U(x)$ with continuous first derivative which interpolates the sample points V_j with a least-mean-square difference, but with the constraint that the derivatives of $U(x)$ are not allowed to fluctuate excessively to obtain the least noisy estimate of the actual derivatives of the sampled function. This is expressed as the problem of minimizing an energy functional E, defined as the mean square difference between the interpolating function and the samples, subject to a penalty on excessively large second derivatives. The strength of the penalty is controlled by a parameter λ, called the regularization parameter:

$$E = \sum_j \left(U(x = j) - V_j\right)^2 + \lambda \int \left(\frac{d^2 U}{dx^2}\right)^2 dx. \qquad (1)$$

It is shown that the $U(x)$ minimizing E in (1) is obtained by convolving V_j with an almost exactly Gaussian kernel, and the width of this kernel increases with λ. We may use this result by exploiting a fundamental connection between the minimum of an energy functional and the operating point of a circuit. It is known from circuit theory that Kirchhoff's laws and the constituent relations of the components drive a network to a state of minimum energy dissipation, so it is reasonable to construct a network whose energy dissipation is described by (1). The network equations may be obtained directly by setting the derivative of the right-hand side of (1) to zero.

Using a discrete estimate of the second derivative in (1), we get

$$E = \sum_j (U_j - V_j)^2 + \lambda \sum_j (U_{j+1} + U_{j-1} - 2U_j)^2 \qquad (2)$$

where $U_j = U(x = j)$. This is a quadratic form, and therefore has a unique minimum where $\partial E / \partial U_j = 0$ for all j, so

$$0 = 2(U_j - V_j) + \lambda \frac{\partial}{\partial U_j} \sum_i (U_{i+1} + U_{i-1} - 2U_i)^2 \qquad \text{for all } j. \qquad (3)$$
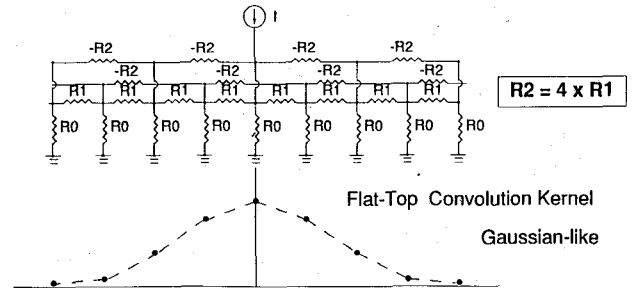

Fig. 2. 1D mesh with negative resistors between second nearest neighbors produces a convolution with a flat top.

Differentiating the terms in the sum and noting that $\partial U_i / \partial U_j = 0$ if $i \neq j$,

$$0 = (U_j - V_j) + \lambda \left(6U_j - 4(U_{j-1} + U_{j+1}) + (U_{j-2} + U_{j+2})\right). \qquad (4)$$

This describes the node equations of a one-dimensional mesh [17] consisting of positive resistors (R_1) connecting nearest-neighbor nodes (i.e., $j - 1, j$ and $j, j + 1$), negative resistors ($-R_2 = -4R_1$) connecting second nearest neighbors, and resistors $R_0 = \lambda R_1$ to ground from every node, which are the leakage resistors described previously in the qualitative model (Fig. 2). The V_j correspond to voltage excitations in series with the leakage resistors. The network will produce as an array of node voltages (U_j) the convolution of the array of excitation voltages (V_j) with a Gaussian kernel whose width is controlled by λ. If $\{V_j\}$ were a set of photovoltages consisting of samples along a scan line through an image, the output set of voltages produced by the network would be the smoothed scan line.

The desired smoothing in an image, however, must take place across two dimensions. To obtain this, samples of a 2D image as a matrix of photovoltages should drive a *two-dimensional* mesh to obtain the desired result. The one-dimensional prototype of a Gaussian convolution mesh must then be extended to implement the kernel with circular symmetry in two dimensions. Noting, for instance, that a two-dimensional Gaussian function $G(x, y)$ is separable, that is, $G(x, y) = G(x) \cdot G(y)$, the desired 2D convolution may be obtained by driving an array of 1D meshes parallel to the y axis with the matrix of sampled photovoltages, and an identical array of 1D meshes along the x axis with the matrix of *buffered* outputs from the first array. This is not very efficient in hardware, because each mesh must have independent active circuits to produce the negative resistances, and an intermesh buffer must be used at every node.

Another possible implementation on a 2D rectangular grid is to connect every node to its *four* nearest neighbors oriented 90° apart with resistors R_1, and the *four* second nearest neighbors at the same orientations with resistors $-R_2$. The simulated spatial impulse response of this network decayed more rapidly along the diagonals than axially, producing an unacceptably large deviation from

(a)

● **2D Gaussian-like**

● **Good Circular Symmetry**
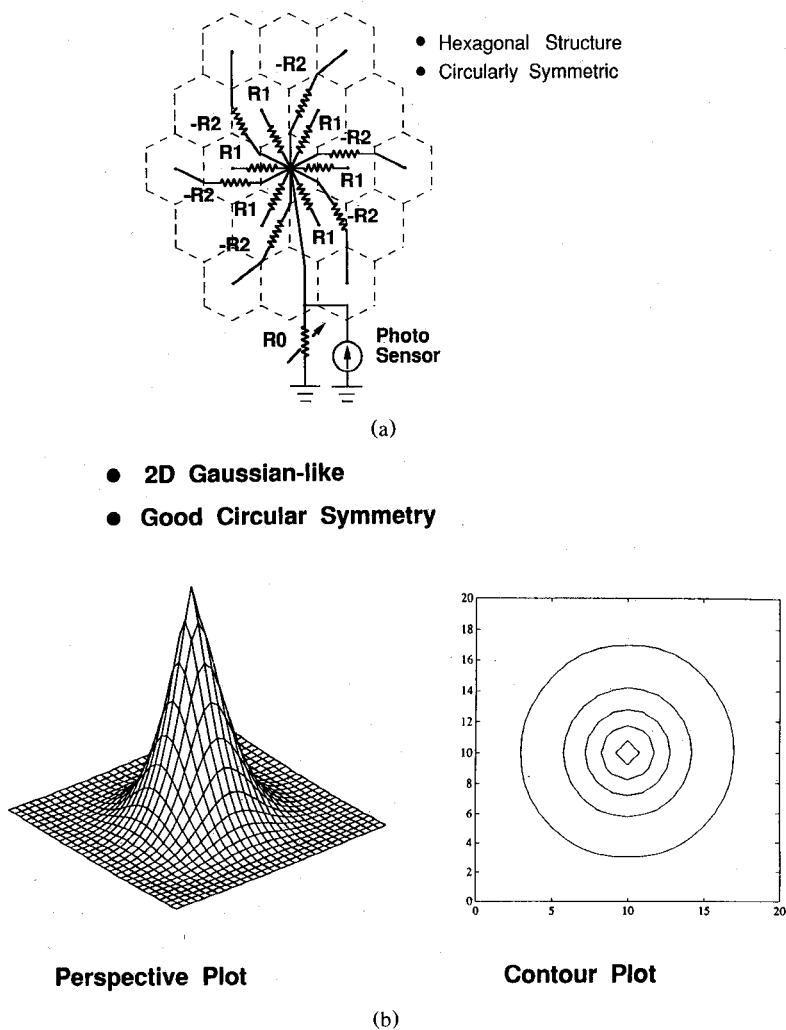

Perspective Plot   **Contour Plot**

(b)

Fig. 3.   (a) Extension of the mesh to 2D on a hexagonal grid produces (b) the best circular symmetry in the convolution kernel.

circular symmetry. A better circular symmetry was obtained by adding similar positive and negative resistive connections along the four diagonal directions, but weighted four times larger in magnitude. It became evident that a large number of components would be required to contrive circular symmetry on a rectangular grid, but not so on a hexagonal grid which inherently possesses a circular symmetry. The image must also be sampled on a hexagonal grid for compatibility with the mesh, which now consists of equal resistive connections 60° apart in orientation to nearest and second nearest neighbors. A hexagonal grid affords the greatest spatial sampling efficiency in the sense that the least photoreceptor sites will attain a desired coverage of the image [18], and the fewest network elements will yield the desired circular symmetry (Fig. 3(a)). The latter was verified in the simulated convolution kernel of this 2D network (Fig. 3(b)).

We required the kernel width to be variable by a factor of 2 under user control. That the convolution width depends on the ratio R_0 / R_1 was known from the synthesis procedure, but the strength of this dependence was

not. Simulations of the network showed a weak dependence (Fig. 4)

$$\text{Convolution width} \propto \left(\frac{R_0}{R_1} \right)^{1/4} . \qquad (5)$$

It was simplest in terms of implementation to keep R_1 and R_2 fixed to preserve the Gaussian shape, and make R_0 alone variable by 16:1 to obtain the desired 2:1 variation in smoothing width.

Several aspects of this design procedure and simulated results invite analysis. Is there a systematic way to generalize a 1D mesh prototype with circular symmetry to 2D? Is the characteristic function of this combination of positive and negative resistors stable in space (i.e., does it decay rather than oscillate indefinitely)? Stable in time? Can the network be generalized to other convolution functions? What is the analytical relation between the width of the convolution function and the network elements? We have answered some of these questions elsewhere [19].
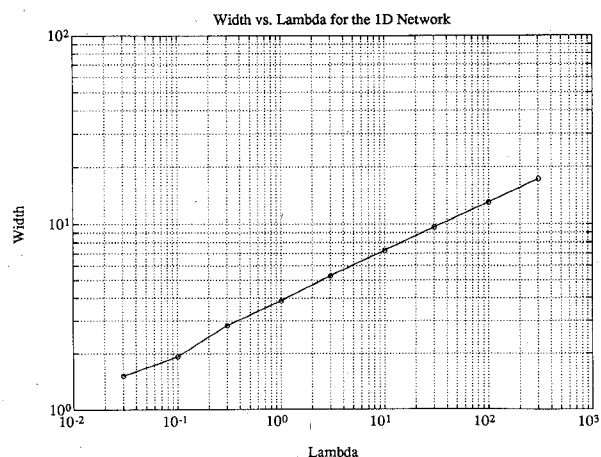
Fig. 4. The width of the convolution kernel increases as the 1/4th power of the grounded resistor.

III. CIRCUIT DESIGN

The practicality of implementing this signal processing technique depends greatly on whether it is realizable on a standard (digital) CMOS IC process. We discuss now the circuit design of the required components, including the photosensors, and the special considerations for layout of this highly interconnected 2D network as a monolithic integrated circuit.

A. Logarithmic Photoreceptor

An image focused on the chip surface may be sampled by a matrix of photoreceptors, one at every node of the network. The intensity across a simple image may vary by two to three orders of magnitude in a laboratory environment, more in natural backgrounds, so a linear photoreceptor, which converts the intensity to a proportional voltage or current, would drive the active circuits in the network into saturation. A logarithmic photoreceptor is therefore required, and as studies on image processing have shown, perfectly adequate for the task on hand [3]. Photosensing is most economically obtained using the parasitic vertical bipolar in a CMOS well as a phototransistor, whose collector current becomes proportional to the light intensity incident on the collector junction along the well boundary. This may be compressed into a logarithmic voltage by a diode-connected MOSFET biased in the subthreshold region by the small photocurrent density produced under room lighting conditions. A compact logarithmic photoreceptor is in this way obtained with a two-transistor circuit [20], [21] (Fig. 5).

Although the stimulus to the prototype network in the discussion above was a voltage source in series with the variable resistor R_0, the circuits for the photosensor output and R_0 (described below) are naturally grounded on one end, so the Norton transformation must be invoked to convert the stimulus into a parallel combination of a grounded current source and a shunt resistor. A transconductance photoreceptor buffer was used, consisting a
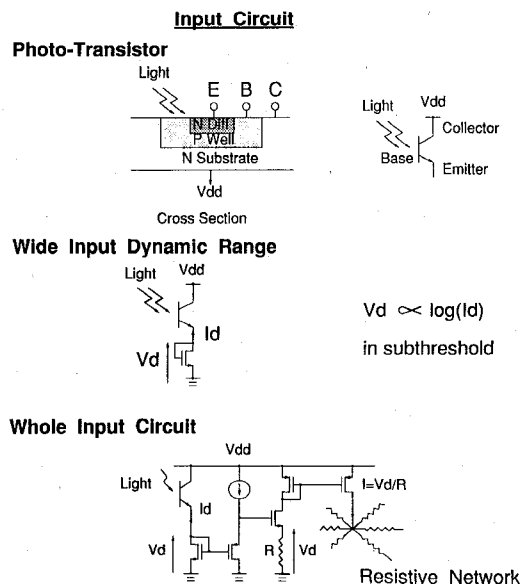

Fig. 5. The vertical bipolar transistor in a CMOS well produces logarithmic compression at the gate voltage by a MOSFET in subthreshold. A transconductance buffer drives the network.

$$I1 = K[(Vc - Vt) Vin - \tfrac{1}{2}Vin^2] \quad \text{(in triode region)}$$

$$I2 = K[(Vc + Vin - Vt) Vin - \tfrac{1}{2} Vin^2]$$

$$I = I1 + I2 = 2 K [Vc - Vt] Vin$$

> $I = Vin / R0$
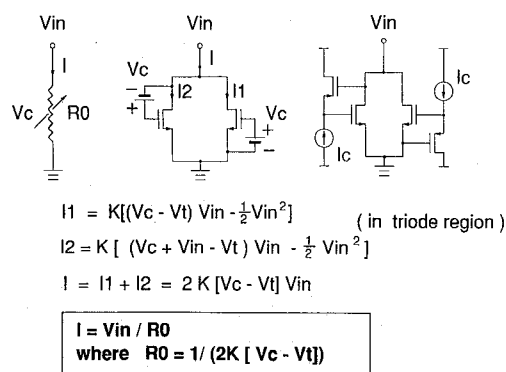> where $R0 = 1/ (2K [Vc - Vt])$

Fig. 6. The linearized variable resistor, with implementation of gate bias.

level-shift PMOS driving a resistively degenerated NMOSFET, which appears to the photoreceptor as a voltage-controlled current source (Fig. 5).

B. Variable Resistor

The width of the convolution kernel is set by a resistor R_0, whose value should ideally be continuously variable under user control. A single MOSFET operating in triode region used as a variable resistor would introduce an undesirable parabolic nonlinearity in the $I-V$ characteristics. Two MOSFET's in parallel obeying the simplified square law equations, however, can exactly cancel each other's parabolic nonlinearity in the triode region of operation if their gate biases are applied in a particular way, and the resulting linearized resistance is controlled by the bias. We used this as the variable resistor (Fig. 6). The floating-gate bias voltages were obtained as the V_{GS} of source-follower FET's carrying a control current.
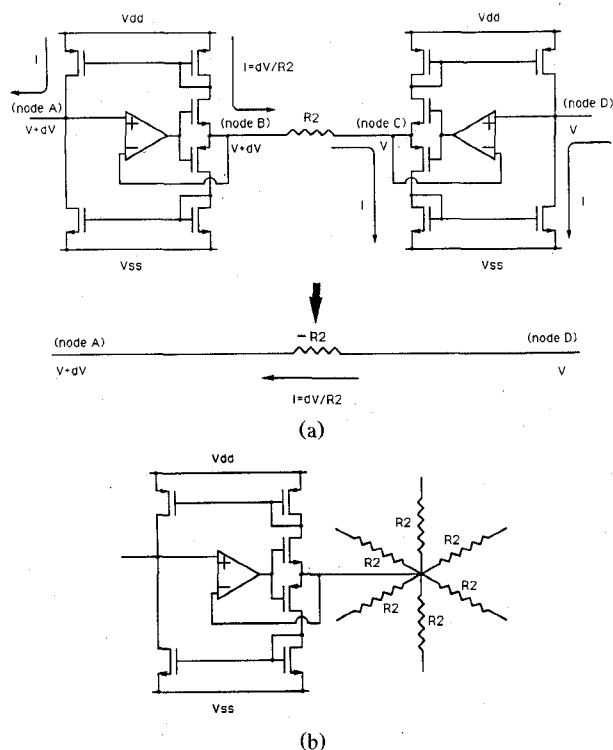
Fig. 7. (a) An NIC inverts the polarity of a resistor. (b) One NIC serves all resistors converging on a node.

The mean network voltage at a given level of photosensor illumination will change with the convolution width: for example, when the convolution width is decreased by making all R_0 large, the mean voltage will also increase because the buffered photocurrents will flow into larger resistors. This will impose the unnecessary demand of a large common-mode range of operation in active circuits such as R_0. We used a scheme to normalize the network inputs by slaving the buffer transconductance of the logarithmic photoreceptor proportionally to R_0, so as to maintain a constant mean network voltage at all illuminations.

C. Network Resistors

The 5-kΩ resistors for the nearest-neighbor internode connections in the network were implemented using p-well diffusions. A Gaussian convolution kernel would be obtained in spite of tolerances in the p-well resistivity as long as the relative magnitude of the positive and negative resistors remains 1:4. To make this ratio on the chip depend only on geometry, both R_1 and R_2 were implemented in the same material, p-well diffusion, and a negative impedance converter (NIC) was attached to R_2 to invert its polarity.

Our NIC implementation (Fig. 7) consists of the combination of a voltage follower and current inverter. The op-amp-based followers at each end of R_2 impose across it the potential difference at their inputs, and the resulting current flow, forced through the Class-B type output

stages, is sourced from or sunk into the positive or negative power supply. Current mirrors in series then apply the same current at the input leads of the followers, inverting the sense of current flow as perceived at the network nodes. A negative resistance $-R_2$ is presented to the network.

Six negative resistors converge on every node in this hexagonal mesh. Six different NIC's are, however, not required at each node; instead, a single NIC placed at the node after the confluence of the resistors will simultaneously make them all negative (Fig. 7(b)). The dc gain in a simple five-FET op amp was large enough to obtain accurate inversion of the resistor $I-V$ characteristics and eliminate the crossover nonlinearity in the Class-B stage. The NIC at every node thus contained only 11 FET's.

D. Layout Considerations

A key concern in the implementation of this network as an IC is whether the usual two layers of metal and one of polysilicon can implement the starlike fan-out of interconnections emanating from every node. We proved to ourselves at the outset of this work that this was possible. A hexagonal grid was obtained by horizontally staggering successive rows of cells, and their interconnections implemented on a Manhattan geometry (Fig. 8(a)). All three available layers of interconnect were used to create abuttable cells. The power, ground, control, and output rails ran parallel to these rows from edge to edge of the chip.

A unit cell, including its portion of interconnect, measured 170×200 μm in 2-μm CMOS (Fig. 8(b)). The area of the photoreceptor collector–base junction, the blank rectangle in the cell layout at the lower left, measured 56×24 μm. No wires were allowed to traverse the photosensor because metal would absorb the incident light. Parasitic photocurrents generated in the source/drain junctions of other active circuits would have negligible effect on the voltages at the low-impedance nodes there. We observe finally that the active circuits occupied only 57% of the cell area, a measure of the toll exacted by the richness of interconnect in this circuit.

E. Output Means

This convolution network accepts a 2D input in the form of an incident image, does 2D signal processing across the resistive mesh, but on a standard IC is restricted to 1D output at the pins along the periphery. The output therefore must be read at the pins (Fig. 9) by accessing one row of nodes at a time, and, at least in this implementation, becomes the bottleneck to the throughput rate. Addressable MOS switches were used to connect every node to output lines, and on-chip vertical bipolar transistors connected as emitter followers served as analog buffers at the pads. The speed of signal processing was determined by the relaxation time of this unclocked network, but a clock was introduced at the output to scan out the rows. To relieve this bottleneck, one can

• **Double Metal, Single Polysilicon**
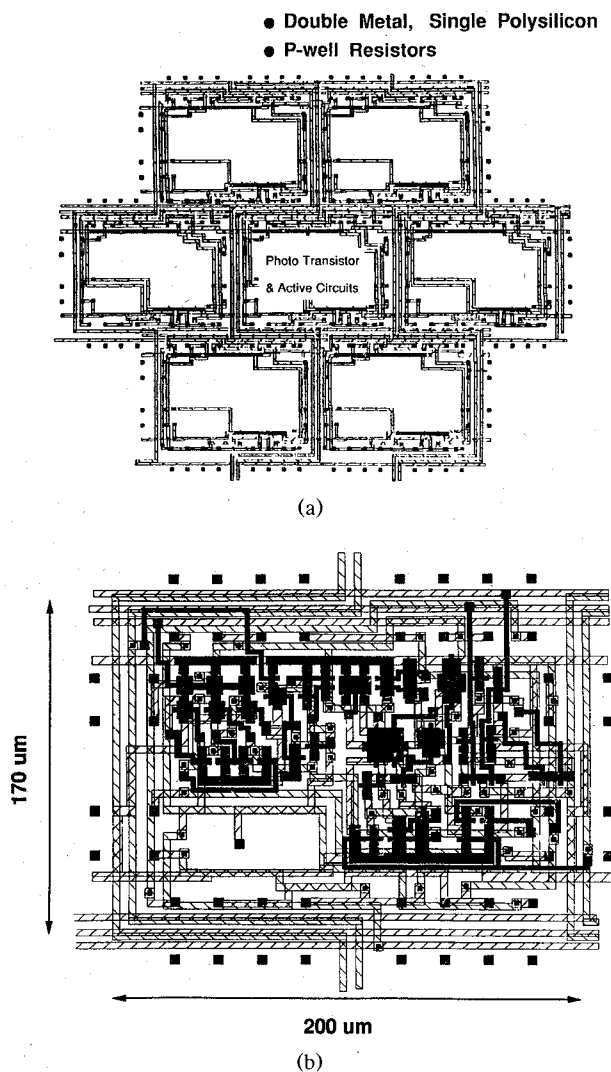• **P-well Resistors**

(a)

(b)

Fig. 8. (a) The layout of interconnects among a cluster of seven cells on a hexagonal grid; the blank areas contain the photoreceptor and associated active circuits in each cell. (b) Unit cell layout.
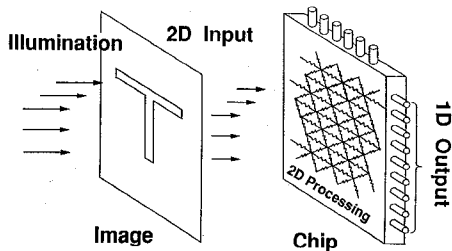

Fig. 9. Output mechanism. The network has 2D input, accomplishes 2D signal processing, but is forced to output results in 1D.

envisage connecting several 2D computational IC's performing a cascade of low-level vision tasks, with micro solder balls joining together matrices of pads on their surfaces, or through via holes on the back sides of the chips. This technique, originally developed for "flip-chip" mounting, is used at very high densities today to mate 2D focal plane array sensors to active substrates [22]. Once the desired data reduction has taken place at the output
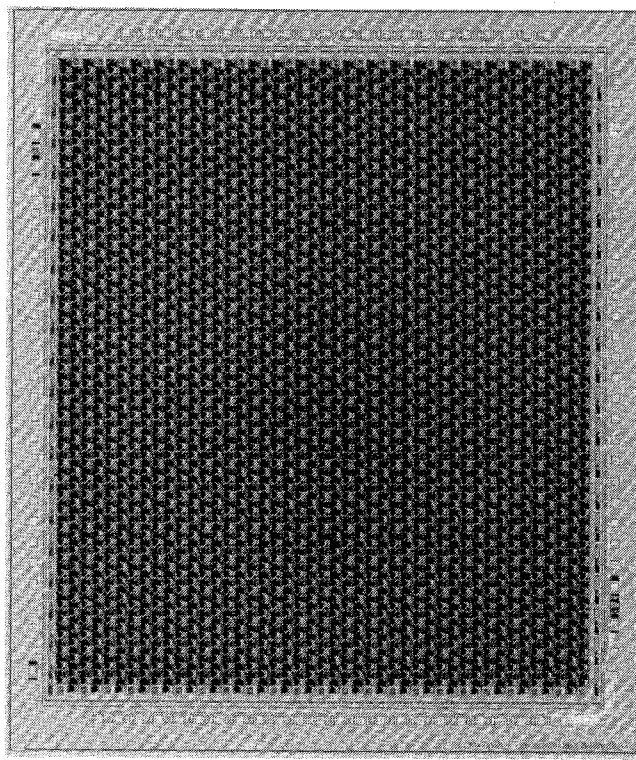

Fig. 10. Chip photograph.

of the such a cascade of chips, a few high-level outputs containing image features could be scanned out in parallel on pins with no loss in throughput speed.

IV. EXPERIMENTAL RESULTS

We were able to fit a 45×40 array of unit cells on a 7.9×9.2-mm die, the largest die size available to us through the MOSIS foundry service. Power supplies of $+5$ and -5 V were used, mainly for convenience in circuit design; the circuits could be modified with a minor effort for operation on a single 5-V supply. The fabricated chip (Fig. 10) contained more than a 100 000 transistors and was fully functional.

The network response to optical input was measured by shining light on the exposed chip, and reading the outputs using a specially developed interface board under control of a personal computer. An array of analog column voltages along an addressed row were digitized and stored, and the smoothed output image reconstructed on the computer screen after all rows had been scanned.
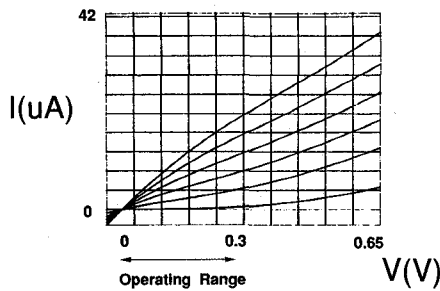
A. Component Characteristics

Test circuits were included to independently verify operation of some of the key building blocks in the network. The log compression FET and the transconductance buffer following the photosensor gave the desired log-linear relationship across 2.5 decades of photocurrent (Fig. 11(a)). The variable resistor could be changed by the control current by a factor of $16:1$ in magnitude, from 20 to 320 kΩ (Fig. 11(b)). The network simulations described
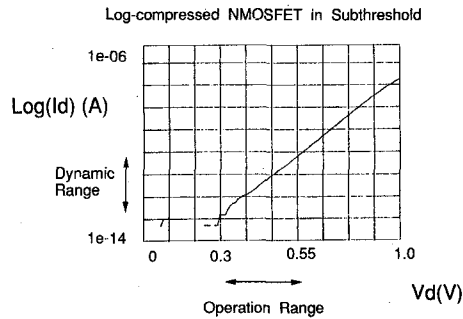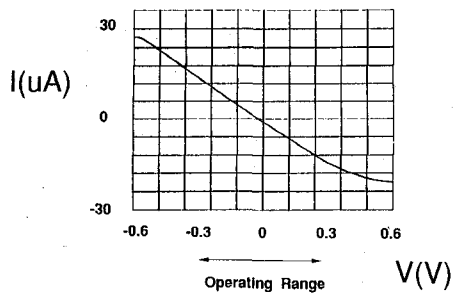
Input Dynamic Range = 2.5 Decades

(a)

R0 = 20 k ohm - 320 k ohm
a factor of 16

(b)

R2 = -20 k ohm

(c)

Fig. 11. Measured characteristics of the component circuits: (a) logarithmic compression at the photoreceptor output (V_d) versus photocurrent; (b) the variable resistor, which becomes nonlinear when one FET goes from triode to saturation; and (c) the negative resistor.

previously predict that this would yield the desired 2:1 variation in convolution width. A strong nonlinearity in the $I-V$ characteristics appeared for voltages larger than 0.3 V, but we had designed the range of the network voltages not to exceed this value under normal illumination. A negative resistor of the desired value was also obtained (Fig. 11(c)), with very little observable nonlinearity at applied voltages of 0.3 V of either polarity.

B. Response to Optical Inputs

The network function was characterized with two simple incident images, a pinhole excitation representing a spatial impulse, and the character "T." The images were produced on the chip surface by light transmitted through
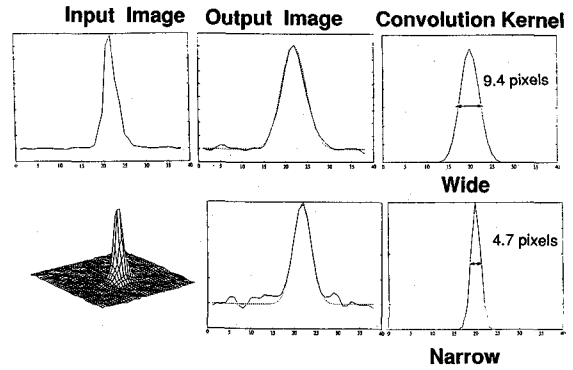

Fig. 12. Measured convolution kernel of the network. The measured network stimulus is deconvolved from the output. Dashed lines superimposed on output show the numerical smoothing used.
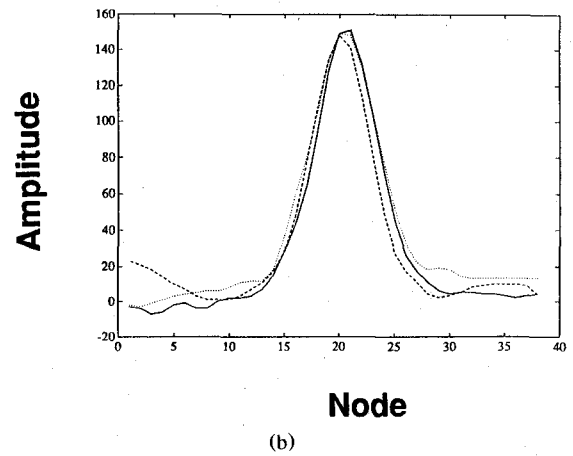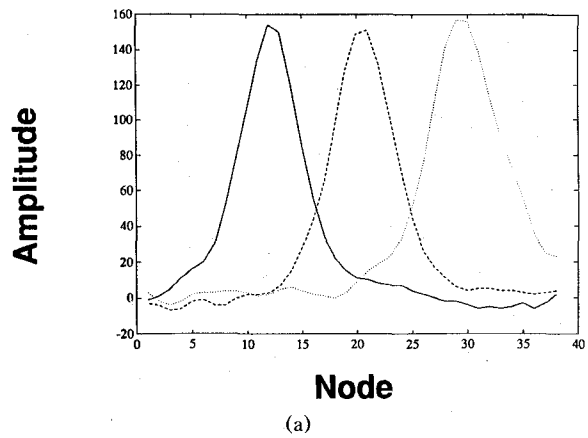

(a)

(b)

Fig. 13. The uniformity of output (a) across one chip, and (b) between three chips.

a mask used in place of the lid on the cavity of the ceramic PGA package. We had also made provision on the IC to measure the actual compressed signal driving the network, so that the true network function could be obtained by deconvolving it from the measured output.

The convolution kernel was thus deduced from measurements of the network input and output (Fig. 12). It was difficult at this sampling resolution to accurately
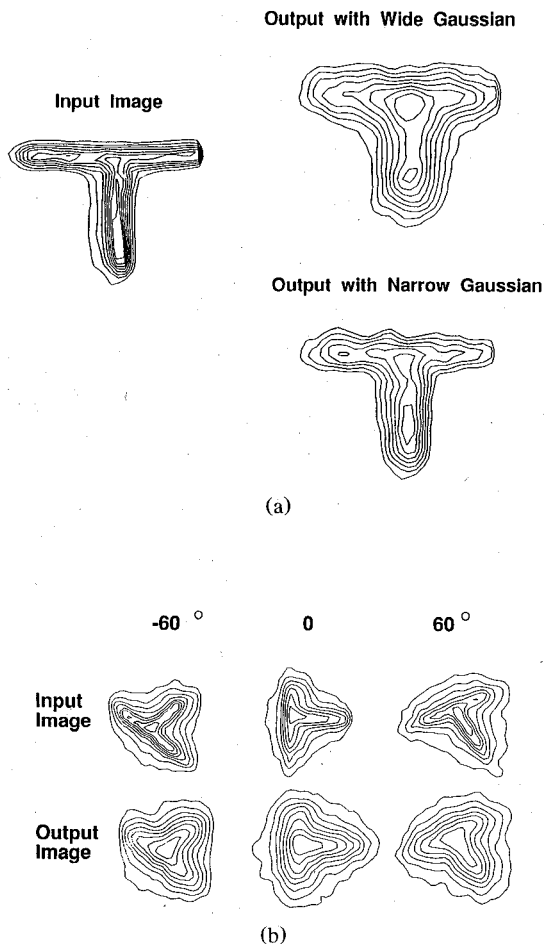
Output with Wide Gaussian

Input Image

Output with Narrow Gaussian

(a)

-60 °   0   60 °

Input
Image

Output
Image

(b)

Fig. 14. (a) Measured outputs at two different smoothing widths on character "T." (b) Uniformity of network action versus rotation.

(a)

node(V)

Time(sec)

(b)

Fig. 15. (a) An 8×8 subnetwork simulated at the transistor level on SPICE, and (b) at various distances away from excitation, showing settling within 2 μs.

ascertain that it was a Gaussian function, but the characteristic inflection in the function as it approaches the peak value was evident. This would not appear unless the network contained negative resistors. We were able to change the full width at half maximum of the kernel by a factor of 2, from 4.7 to 9.4 pixels wide, by changing R_0 across its full span with the control current. The network output was most noisy at its tails at minimum R_0, and we had to use smoothing in the sense of a least-mean-square fit to deduce the kernel function. Light through the pinhole nominally sampled only a small neighborhood on the chip; we moved the pinhole to points on the chip either side of the center, and found an acceptable uniformity in the response (Fig. 13(a)), which is determined here by MOSFET matching across the extent of the chip surface [23]. The slight uptilt of the output at the ends of the measured response was caused by the edge effect when the network terminates at the chip boundary. The uniformity across three chips was also acceptable at this sampling resolution (Fig. 13(b)), except for one chip where a particularly large uptilt appears.

The smoothing effected by the network on a character "T" was also measured (Fig. 14(a)), and its symmetry after rotations relative to the chip axis verified (Fig. 14(b)). Both were satisfactory.
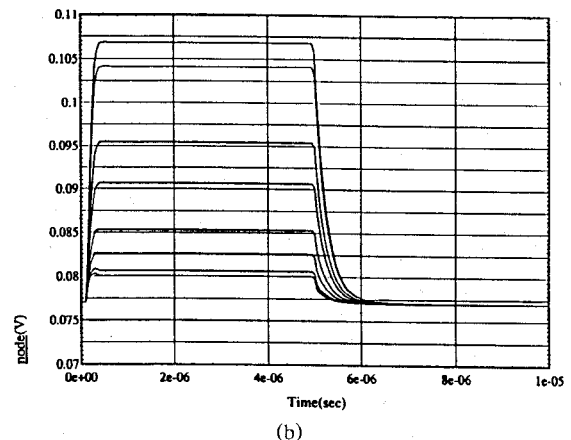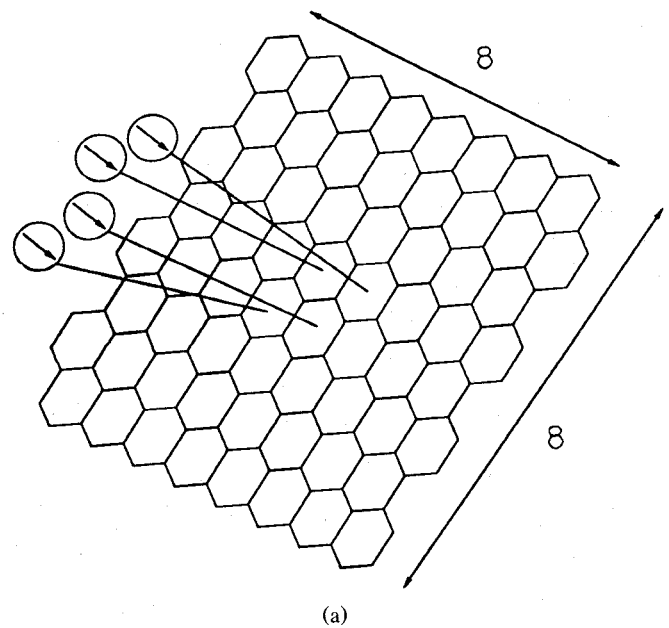
Precautions were required in making the measurement to compensate for the effects of the 2-W power dissipation when no heat sink was mounted on the package. This large power dissipation produced a thermal gradient across the IC, peaked at the center with circularly symmetric isotherms spreading out towards the chip boundary. We deduced this from a corresponding pattern in photoreceptor dark currents, which appeared as a stimulus to the network in the absence of an optical input. This had to be calibrated and subtracted from all measurements to obtain the true optical response. We emphasize that this relatively large power dissipation was not fundamental to the network; 75% of it was due to an unnecessarily large bias current in one building block, the control circuit for the variable resistor. A further reduction in quiescent power could be obtained by devising a voltage drive to the network nodes, because the current sources in the present implementation produce some steady power dissipation through R_0, even when the chip is not illumi-

TABLE I
ELECTRICAL CHARACTERISTICS

Photosensor sites	45×40
Sampling geometry	Hexagonal
Area per pixel	170×200 μm
Rise time of network (10–90%)	2 μs
Rise time of photosensors	20 μs
Width of convolution (FWHM)	4.7–9.4 pixels
Chip size	7.9×9.2 mm
Technology	2-μm CMOS, single poly, double metal
Power dissipation	2 W (75% in one function block)

nated. The power dissipation could be made even smaller by scaling down all the currents in the IC, but at a trade-off of longer relaxation times.

The settling time of the entire network in response to a step input from the photoreceptors determined the 2D computational speed. For all practical purposes, a step change in a photoreceptor has only to propagate a few nodes away before the decay in the convolution function will swamp it out, and the voltages at nodes farther away will remain relatively unchanged. We simulated an 8×8 subnetwork at the transistor level on SPICE, and the results indicated settling in less than 2 μs in response to a step in photocurrent (Fig. 15). However, a settling time of 20 μs was experimentally observed in response to illumination from a light chopper, which we surmise was dominated by the slow response of the phototransistors [20]. The graceful settling in the transient SPICE simulation verified the stability of the network response in time. A similar waveform of the settling of node voltages was also observed experimentally.

The electrical performance of the Gaussian convolution IC is summarized in Table I.

V. CONCLUSIONS

Parallel processing of images per pixel will offer the highest possible speed in functions related to low-level vision. This is indeed the present trend in real-time hardware for digital image processing. We have described a single-chip analog implementation of this concept to perform a Gaussian convolution with the use of an active mesh. Although it may be argued that a variable focus lens also effects this function, there are two significant differences: the active resistive mesh may be extended to many different convolution functions, including orientation selective ones [19], most of which cannot be simply implemented with geometric optics; furthermore, no mechanical system could attain the physical compactness and microsecond control of the convolution functions. The difference in output of two independent meshes on the same chip, for example, could implement the much sought after difference of Gaussian function in image processing [3]. In short, the notion of an active mesh opens many new opportunities for realizing application-specific analog signal processors. Digital signal processors have as advantages an immunity to component noise and mismatches, more ready programmability, and shorter development

times, but tend to be considerably larger chips than their analog equivalents. On the other hand, inaccuracies in analog computation may not be limitations in low-level vision functions, but much more of a detriment in high-level classification tasks. This leads us to believe that compact hardware with the least power dissipation to implement real-time image recognition and classification may ultimately consist of a judicious mix of analog computation of the type described here, and conventional digital signal processing.

REFERENCES

[1] P. A. Ruetz and R. W. Brodersen, "Architectures and design techniques for real time image processing ICs," IEEE J. Solid-State Circuits, vol. SC-22, pp. 233–250, Apr. 1987.
[2] J. Dowling, The Retina: An Approachable Part of the Brain. Cambridge, MA: Harvard University Press, 1987.
[3] D. Marr, Vision. San Francisco, CA: W. H. Freeman, 1982.
[4] C. A. Mead and M. A. Mahowald, "A silicon model of early visual processing," Neural Networks, vol. 1, pp. 91–97, 1988.
[5] E. A. Vittoz, "Future of analog in the VLSI environment," in Proc. ISCAS (New Orleans, LA), May 1990, pp. 1372–1375.
[6] B. Gilbert, "Translinear circuits: A proposed classification," Electron. Lett., vol. 11, pp. 14–16, 1975.
[7] B. Gilbert, "A monolithic 16 channel analog array normalizer," IEEE J. Solid-State Circuits, vol. SC-19, pp. 954–963, Dec. 1984.
[8] H. Kobayashi, J. L. White, and A. A. Abidi, "An analog CMOS network for Gaussian convolution with embedded image sensing," in ISSCC Dig. Tech. Papers (San Francisco, CA), Feb. 1990, pp. 216–217.
[9] T. Poggio, H. Voorhees, and A. Yuille, "A regularized solution to edge detection," Mass. Inst. Technology, Cambridge, MA, AI Memo, May 1985.
[10] T. Poggio, V. Torre, and C. Koch, "Computational vision and regularization theory," Nature, vol. 317, pp. 314–319, Sept. 1985.
[11] J. Babaud, A. P. Witkin, M. Baudin, and R. O. Duda, "Uniqueness of the Gaussian kernel for scale-space filtering," IEEE Trans. Pattern Anal. and Mach. Intell., vol. PAMI-8, pp. 26–33, Jan. 1986.
[12] T. K. Hogan, "A general experimental solution of Poisson's equation for two independent variables," J. Inst. Eng. (Australia), vol. 15, pp. 89–92, Apr. 1943.
[13] G. Liebmann, "Solution of partial differential equations with a resistance network analogue," Brit. J. Appl. Phys., vol. 1, pp. 92–103, Apr. 1950.
[14] G. W. Swenson, Jr. and T. J. Higgins, "A direct current network analyzer for solving wave equation boundary value problems," J. Appl. Phys., vol. 23, pp. 126–131, Jan. 1952.
[15] J. R. Hechtel and J. A. Seeger, "Accuracy and limitations of the resistor network used for solving Laplace's and Poisson's equations," Proc. IRE, vol. 49, pp. 933–940, May 1961.
[16] C. A. Mead, Analog VLSI and Neural Systems. Reading, MA: Addison Wesley, 1989.
[17] T. Poggio and C. Koch, "Ill-posed problems in early vision: From computational theory to analogue networks," Proc. Roy. Soc. London, vol. B-226, pp. 303–323, 1985.

[18] D. Dudgeon and R. Mersereau, *Multidimensional Signal Processing*. Englewood Cliffs, NJ: Prentice Hall, 1984.

[19] J. L. White and A. A. Abidi, "Analysis and design of parallel analog computational networks," in *Proc. Int. Symp. Circuits Syst.* (Portland, OR), June 1989, pp. 70–73.

[20] S. G. Chamberlain and J. P. Y. Lee, "A novel wide dynamic range silicon photodetector and linear imaging array," *IEEE J. Solid-State Circuits*, vol. SC-19, pp. 41–48, Feb. 1984.

[21] C. Mead, "A sensitive electronic photoreceptor," in *Proc. 1985 Chapel Hill Conf. VLSI* (Chapel Hill, NC), 1985, pp. 463–471.

[22] S. B. Stetson, D. B. Reynolds, M. G. Stapelbroek, and R. L. Stermer, "Design and performance of blocked impurity band detector focal plane arrays," in *Proc. SPIE*, vol. 686 (San Diego, CA), Aug. 1986, pp. 48–65.

[23] M. J. M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers, "Matching properties of MOS transistors," *IEEE J. Solid-State Circuits*, vol. 24, no. 5, pp. 1433–1440, Oct. 1989.

Haruo Kobayashi was born in Utsunomiya, Japan, in 1958. He received the B.S. and M.S. degrees in information physics and mathematical engineering from the University of Tokyo, Tokyo, Japan, in 1980 and 1982, respectively. From 1987 to 1989 he was at the University of California, Los Angeles, where he received the M.S. degree in electrical engineering in 1989.

He joined Yokogawa Electric Corporation, Tokyo, Japan, in 1982, where he has been engaged in the research and development of an FFT analyzer, a mini-supercomputer, and an LSI tester.

Mr. Kobayashi is a member of the Institute of Electronics, Information and Communication Engineers of Japan and the Society of Instrument and Control Engineers of Japan.

Joseph L. White (S'88) received the B.S. degree in applied physics from the California Institute of Technology, Pasadena, in 1982, and the M.S. and Ph.D. degrees in electrical engineering from the University of California, Los Angeles, in 1983 and 1991, respectively.

He has previously worked for the Hughes Aircraft Space and Communications Group and the Rand Corporation. His research interests include image processing and computer vision.

Asad A. Abidi (S'75–M'81) was born in 1956. He received the B.Sc.(Hon.) degree from Imperial College, London, in 1976 and the M.S. and Ph.D. degrees in electrical engineering from the University of California, Berkeley, in 1978 and 1981, respectively.

He was at Bell Laboratories, Murray Hill, NJ, from 1981 to 1984 as a Member of the Technical Staff in the Advanced LSI Development Laboratory. Since 1985 he has been at the Electrical Engineering Department of the University of California, Los Angeles, where he is an Associate Professor. He was a Visiting Faculty Researcher at Hewlett Packard Laboratories during 1989. His research interests are in high-speed analog integrated circuit design, parallel analog signal processing techniques, device modeling, and nonlinear circuit phenomena.

Dr. Abidi served as the Program Secretary for the International Solid-State Circuits Conference from 1984 to 1990, and is presently associated with the Symposium on VLSI Circuits, with the IEEE Solid-State Circuits Council, and as an Associate Editor with the IEEE JOURNAL OF SOLID-STATE CIRCUITS. He received the 1988 TRW Award for Innovative Teaching.