ISCAS 2000 - IEEE International Symposium on Circuits and Systems, May 28-31, 2000, Geneva, Switzerland

CMOS DYNAMIC LINKING NETWORKS FOR REAL-TIME HUMAN FACE TRACKING

R. G. Spencer

Analog & Mixed-Signal Center
Texas A&M University
College Station, Texas 77843-3128

ABSTRACT

Dynamic linking networks are very effective in locating human faces, as they are invariant to changes in position, orientation, scale, deformation, and partial lighting. Additionally, these networks are well suited for analog VLSI using very simple CMOS circuits such as programmable ring oscillators, strongly coupled diffusion lattices, and AC-coupled Hebbian synapses. Dynamic linking networks are orders of magnitude faster than digital computers, thus they provide a fast, efficient, and robust means of tracking faces in real-time. Transistor-level simulation results are presented for the .5μm TSMC CMOS technology with 3-volt supply.

1. INTRODUCTION

Recognizing faces in real-time is a challenging task, due in part to the time required to extract features. This processing time is often directly related to the size of the image. Unfortunately, a large part of the image typically contains irrelevant information that distracts recognition processes from regions of interest (ROI). Significant improvements can be made in processing time and error rates by finding and focusing attention on these regions only. Hence, there is a great need for fast and efficient ROI locating processes.

Some progress has been made in ROI tracking using techniques such as morphological processing, sub-sampling, spline-wavelet preprocessing, wavelet transforms, log-polar mapping, and multilayer perceptron networks [1], [2]. Although these techniques can be made to work in well-controlled environments, they are generally not robust in a wide range of real-world variations such as changes in orientation, scale, deformations, and lighting. Furthermore, many of these techniques are relatively complex, making them poor candidates for VLSI.

Fortunately, *dynamic linking networks* offer a straightforward, robust approach that is simple enough to integrate on-chip using ring oscillators, *strongly coupled diffusion lattices*, and *AC-coupled Hebbian synapses*. Because diffusion lattices can calculate many local summations of oscillator activity simultaneously, they are orders of magnitude faster than digital computers. As a whole, dynamic linking networks constitute a very fast, efficient, and robust means of real-time ROI tracking.

2. DYNAMIC LINKING NETWORKS

One of the most general descriptions of a face is a list of its features without any information as to how they are arranged spatially. Such a description is a simple *combination* of features. On the other hand, *permutation* pertains to how such features are arranged spatially and thus represents a more *specific* description. Without compensation, a permutation-based description is far too rigid to be invariant to translation, rotation, scale, and deformation. Fortunately, most faces have a rather distinct relative distribution (combination) of pixel intensities, making them identifiable against cluttered backgrounds and other objects while maintaining invariance to real-world variations.

So the question is, how can such a high-dimensional combinatorial description be represented in hardware? The answer lies in temporal summation of sinusoids. At the lowest descriptive level, a summation of *DC* pixel intensities is *too* general because the individual components do not retain their identity after summation. But a sum of *sine waves* of different frequencies does preserve the identity of its constituents and thus constitutes a *separable composite*. Hence, a system that constructs such composite signals from sine waves that encode pixel intensity as frequency of oscillation and correlates them with known temporal exemplars such as those obtained from a normalized average of all faces, is able to discriminate between faces and non-faces while still maintaining invariance to real-world variations of the faces themselves. A block diagram of such a system is shown in Fig. 1.

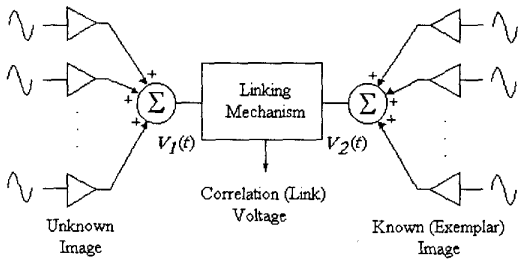

Fig. 1. Dynamic linking network with oscillatory illuminance encoding for finding faces in uncontrolled environments.

The linear combiner on the left side of Fig. 1 produces a composite signal, $V_1(t)$, from a set of oscillatory signals determined by relative reflectances of an unknown image. The linear combiner on the right produces a similar composite signal, $V_2(t)$, from a known (exemplar) image. The linking mechanism in-between computes the correlation between $V_1(t)$ and $V_2(t)$. If the two relative pixel intensity distributions are similar, then the two composite signals will be correlated and the absolute value of the correlation voltage

0-7803-5482-6/99/$10.00 '2000 IEEE

will grow. On the other hand, if the two patterns are different, the correlation voltage will decay to zero.

Finding regions of interest in real images requires that many *local* summations and linking mechanisms be distributed around the unknown image. Then the links that become greater than a certain threshold can be used as a binary mask to determine where feature extraction and classification techniques should be applied. Although such a distributed architecture can be implemented on a digital computer, the shear size and number of summations that must be computed requires large amounts of computational resources, making the software implementation impractical. Hence, there is a need for hardware implementation using analog VLSI.

3. HARDWARE IMPLEMENTATION

Dynamic linking networks are well suited for analog VLSI. Ring oscillators with programmable capacitor banks and cascode current-output stages can be used to encode pixel intensity. Summation, which typically takes a great deal of computational resources on a digital computer, can be implemented by *strongly coupled diffusion lattices*. And the linking mechanisms can be implemented by *AC-coupled, transconductance multipliers*.

3.1. Diffusion Lattices

Diffusion lattices like that shown in Fig. 2 can be used to calculate many local summations of oscillator activity that drive the adaptation in a dynamic linking network. As a result, there is no need to include a linear combiner for every pixel location; a substantial savings in processing time and area! As long as the lattice nodes are strongly coupled; i.e. lateral resistances, R_{lat}, are small compared to the vertical resistances, R_{ver}, and the oscillators inject *currents* into their respective nodes, then the voltage that develops at each node is an approximate local average of neighboring frequencies.

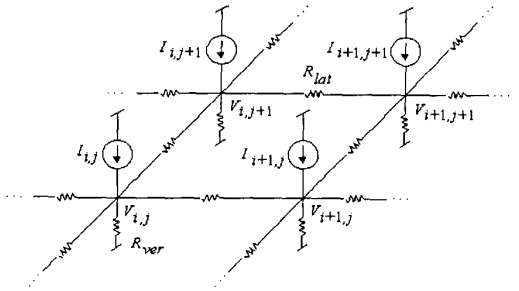

Fig. 2. A strongly coupled diffusive averaging lattice.

Because the lateral resistors must be small, they can be implemented with serpentine segments of polysilicon, but the vertical resistors must be implemented with active components to conserve area. A simple, yet fairly linear approach is to use a complementary shunted transistor pair as shown in Fig. 3. The linearity of this scheme is good in the voltage range of interest (0 to .8V) when (1) is satisfied

$$\mu_p \left(\frac{W}{L}\right)_p = \mu_n \left(\frac{W}{L}\right)_n \qquad (1)$$

approximately [3]. The parameters μ_n and μ_p are the electron and "hole" mobilities, respectively, and W and L are the width and length of the transistors. For gate voltages of V_{bn} = 2.5V and V_{bp} = .5V and aspect ratios of .25 and .93, equivalent resistances of approximately 31kΩ are realized.

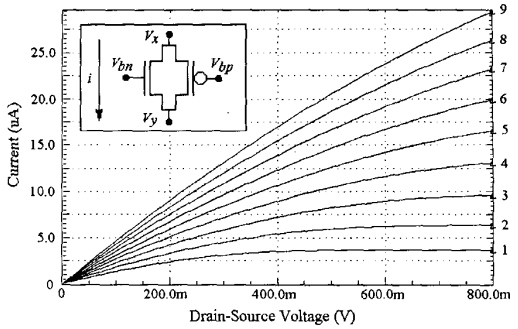

Fig. 3. Linearity of a complementary shunted transistor pair for implementing passive resistors. The control voltages V_{bn} and V_{bp} were swept from (1.4,1.6) to (3,0) in steps of .2V.

3.2. Programmable Ring Oscillators

In order to encode pixel intensity as frequency of oscillation, an oscillator must be allocated for every pixel in the image whose frequency is programmable over a range of at least four bits. A single-ended ring oscillator with programmable capacitor bank and cascode current-output stage is well suited for this purpose (see Fig. 4).

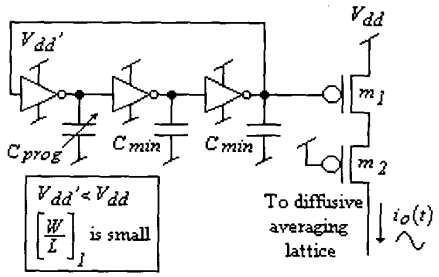

Fig. 4. Programmable ring oscillator with current-output stage. The layout without capacitors is 980 sq. μm.

The cascode output stage provides a current output with a high output resistance. The output resistance must be high in order to minimize intermodulation between oscillators. Each oscillator must inject a current into its respective node in the lattice that is independent of the time-varying voltage that develops at that node. The well-known cascode structure achieves this objective by exploiting one of the many benefits of negative feedback; increased output resistance.

The four-bit capacitor bank, denoted by C_{prog} in Fig. 4, allows the frequency of oscillation to be programmed over a small range as shown in Fig. 5. In order to minimize area, capacitors must be on the order of tenths of picofarads, making the on-capacitance of programming switches non-negligible. As a result, a binary-weighted sizing strategy must be employed to maintain good linearity between the

code and frequency of oscillation. The binary-weighted sizing strategy is contrasted with the uniform sizing strategy in Fig. 5.

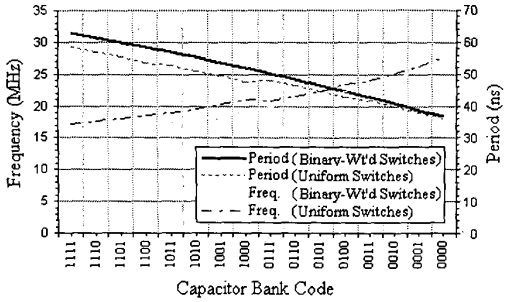

Fig. 5. Linearity of the programmable oscillator with respect to uniform vs. binary-weighted sizing strategies.

3.3. Linking Mechanisms

Two correlation mechanisms that can be used for dynamic linking are the *Widrow-Hoff* [4] and *Hebbian* learning rules [5]. Both rules have certain advantages and disadvantages. The one-dimensional form of the Widrow-Hoff learning rule is given by (2):

$$\frac{dV_o(t)}{dt} = c\left[V_2(t) - V_o(t)V_1(t)\right]V_1(t) \quad (2)$$

where $V_o(t)$ is the output voltage, c is the learning rate, and $V_1(t)$ and $V_2(t)$ are the signals to be correlated. The advantage of this rule is the fact that $V_o(t)$ converges to a high absolute value when $V_1(t)$ and $V_2(t)$ are temporally correlated *without being sensitive to DC offsets*. The disadvantages, however, are the area, power consumption, common-mode rejection, and linearity of the blocks required to implement it.

The Hebbian learning rule, which is the simplest learning rule in neural learning theory, is given by (3):

$$\frac{dV_o(t)}{dt} \doteq cV_1(t)V_2(t) \quad (3)$$

where all variables are the same as in (2). The overwhelming advantage of this rule is simplicity; the only block required to implement it is a transconductance multiplier. The disadvantage, however, is that it is unbounded in the presence of constant correlation [6]. As a result, it is very sensitive to DC offsets which lead to saturation in voltage levels, regardless of the AC components. Fortunately, DC offsets can be removed with *bypass capacitors*, making the Hebbian learning rule extremely practical for VLSI.

The Hebbian synapse can be implemented at the transistor level by the transconductance multiplier shown in Fig. 6. This circuit is essentially a Gilbert cell [7], [8] with active load and current output. The simple addition of a capacitor at the output makes the overall block an integrator. The gain-bandwidth product is approximately 80MHz and its power consumption is 143 microwatts. The layout area without the capacitor is approximately 1280 sq. μm.

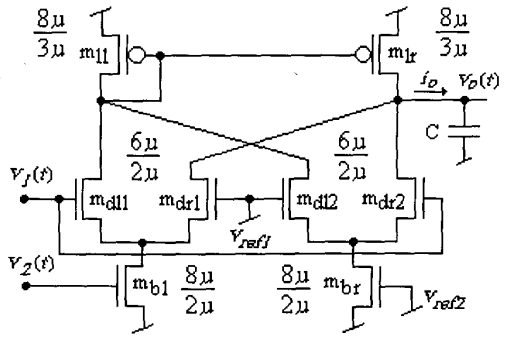

Fig. 6. Transconductance multiplier/integrator used to implement a Hebbian synapse.

4. SIMULATION RESULTS

Several simulations were conducted to characterize the dynamic linking mechanism at the transistor level. The circuit shown in Fig. 7 was used to test the transient behavior of the Hebbian synapse in response to a range of similar frequency distributions. The circuit consisted of two 3 x 3 strongly coupled diffusion lattices and one AC-coupled Hebbian synapse. (Oscillatory current sources and vertical resistors in the diffusion lattices are not shown for the sake of simplicity.)

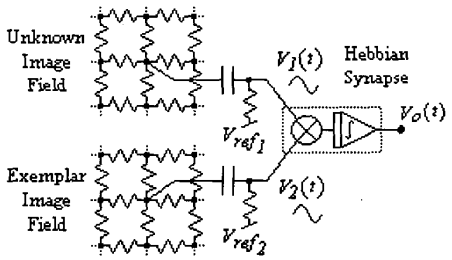

Fig. 7. A small test network for characterizing the AC-coupled Hebbian synapse.

The number of frequency components common to both lattices was swept from zero to nine and the resulting correlation voltages were graphed in Fig. 8. For the case in which the composite signals had no common frequencies, the correlation voltage decayed to 'zero' (approximately equal to

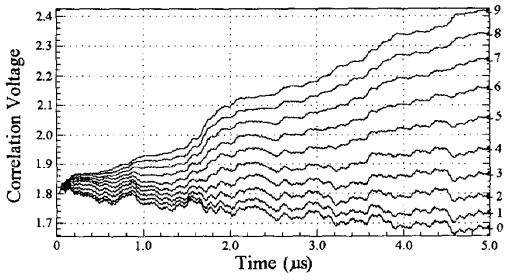

Fig. 8. Correlation voltage versus time for increasing number of common frequency components. (.5μm TSMC)

1.7V). For the case in which all nine frequency components matched, the correlation voltage rose to just over 2.4V in 5 μs. For combinations in between, the correlation voltage rose to a level that was nearly proportional to the number of common frequencies (see Fig. 9).

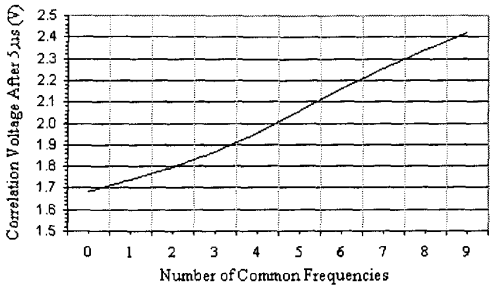

Fig. 9. Correlation voltage versus number of common frequency components between diffusion lattices.

Simulation results with real images were performed and are shown in Fig. 10. These results strongly suggest that the network performed well in discriminating between faces and non-faces and was invariant to rotation, scale, translation, and small degrees of deformation and lighting fluctuations. In every trial, a correlation voltage was obtained at each pixel in the test image by correlating the local composite signal with the composite signal at a node directly over the center of the face in the exemplar lattice (not shown). The two images in Fig. 10a show how the network performed in the presence of simultaneous illumination, scale, and translation variance. The left and right images of Fig. 10a are the unmasked and masked test images, respectively. The masked image was obtained by masking out any pixel whose correlation voltage converged to 2.3 volts or greater (1.7 being the baseline). Fig. 10b shows how the network performed in the presence of simultaneous rotation, scale, and translation variance. Fig. 10c shows how the network performed in the presence of deformation and translation. Fig. 10d, e, and f show how the network performed in the presence of background clutter and different and multiple subjects, respectively.

5. SUMMARY

Dynamic linking circuits were presented that are capable of focusing attention on facial regions of an image regardless of a wide variety of real-world variations such as position, orientation, scale, deformation, and partial lighting changes. These networks were shown to be well suited for VLSI using programmable ring oscillators, diffusion lattices, and AC-coupled Hebbian synapses. Strongly coupled diffusion lattices were shown to be able to calculate many local summations of oscillator activity *in parallel*, which constitutes a *significant* improvement in processing speed over digital computers. Correlation voltages in areas where the relative intensity distribution is similar to that of an exemplar image were shown to grow at a rate that is proportional to the number of common frequencies there. It was shown that the links that grow larger than a certain threshold can be used as a binary mask to determine where feature extraction and classification techniques should be applied. In summary, dynamic linking networks are very powerful, robust, and well suited for analog hardware implementation using very simple CMOS circuits.

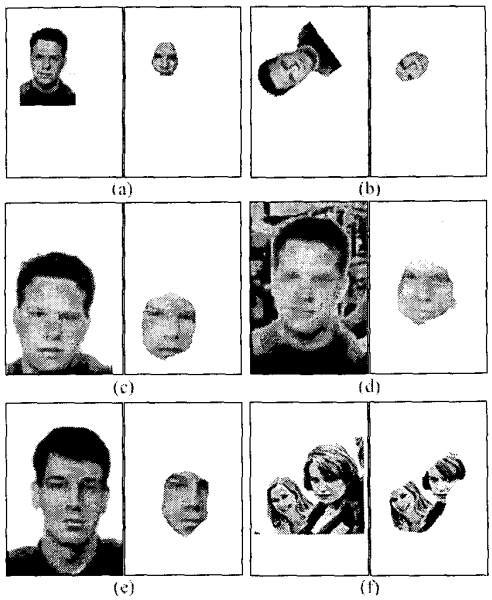

Fig. 10. Simulation results of dynamic linking networks using oscillatory-illuminance encoding for finding faces in uncontrolled environments.

6. REFERENCES

[1] R. Spencer, *Low-Level Neural Circuits and Systems for Artificial Face Recognition*, Ph.D. Dissertation, Dept. of Elect. Eng'g, Texas A&M University, Aug. 1999.

[2] R. Spencer and E. Sanchez-Sinencio, "A monolithic mixed-mode implementation of sum-of-product arrays for performing binary morphological image processing," *IEEE ISCAS '97*, Hong Kong, 1997.

[3] C. Enz, F. Krummenacher, and E. Vittoz, "An analytical MOS transistor model valid in all regions of operation and dedicated to low-voltage and low-current applications," *Analog Integrated Circuits and Signal Processing*, vol. 8, no. 1, pp. 83-114, Jul. 1995.

[4] B. Widrow, "Generalization and Information Storage in Networks of Adaline 'Neurons'," in *Self-Organizing Systems*, M. Jovitz, G. Jacobi, G. Goldstein, Eds., Washington D.C.: Spartan Books, pp. 435-461, 1962.

[5] D.O. Hebb, *The Organization of Behavior*, New York: Wiley, 1949.

[6] G. A. Carpenter and S. Grossberg, *Pattern Recognition by Self-Organizing Neural Networks*, Cambridge, MA: MIT Press, 1991.

[7] B. Gilbert, "A precision four-quadrant multiplier with sub-nanosecond response," IEEE J. Solid-State Circuits, vol. SC-3, pp. 353-365, Dec. 1968.

[8] G. Han and E. Sánchez-Sinencio, "CMOS trans-conductance multipliers: A tutorial," *IEEE Trans. on Circuits and Systems II*, vol. 45, pp. 1550-1563, Dec. 1998.