An Analog Retina Model for Detecting Dim Moving Objects Against a Bright Moving Background

R. M. Searfus, M. E. Colvin, F. H. Eeckman,
J. L. Teeters, and T. S. Axelrod
Lawrence Livermore National Laboratory
Livermore, California 94550

Abstract – **We are interested in applications that require the ability to track a dim target against a bright, moving background. Since the target signal will be less than or comparable to the variations in the background signal intensity, sophisticated techniques must be employed to detect the target. We present an analog retina model that adapts to the motion of the background in order to enhance targets that have a velocity difference with respect to the background. Computer simulation results and our preliminary concept of an analog "Z" focal plane implementation are also presented.**

1 Introduction

We are interested in air and spaceborne surveillance applications that require real-time target detection and tracking against a moving earth background. The scene observed from the surveillance platform may range from a dark earth to bright sunlit clouds and terrain, and the variation in the intensity of a single scene may span several orders of magnitude. As long as the target intensity is sufficiently larger than the variations in the background intensity, simple image processing techniques such as spatial filtering and thresholding can produce satisfactory results; however, for the case of a dim target against a bright, moving background, simple processing methods may produce unacceptable levels of false detections or may completely fail to detect the target.

One approach for reliably detecting and tracking targets under these conditions is to subtract the moving background from the scene, leaving only those objects that have a different velocity than the background. Since the background motion may not be known *a priori* and may change throughout the course of observation, it is important for the sensor to have the capability of adapting to changes in the background velocity. The number of detector signals that must be simultaneously processed[1] imposes a computational demand that exceeds the capability of conventional computer hardware. Furthermore, for a space environment, low-power consumption and compact size are extremely important design constraints.

In this paper, we present a model for an analog retina that adapts to the motion of the background and enhances objects having a velocity difference with respect to the background. A computer simulation of this model is described, and our experience of using the simulation on real and synthetic data is discussed. We also describe a real-time implementation of our model on a PIPE image processing computer, and present a

[1] A minimum detector array of 128x128 pixels is required; an array of 512x512 pixels is desired.

mapping of our model to a "Z" focal plane (Z-plane) technology [?] implementation that addresses the real-time processing requirements and the design constraints for space-based operations.

2 An Analog Retina-like Model

Very sensitive, high-resolution electronic imaging systems exist with capabilities that surpass those of any biological system. However, current electronic imaging systems do not possess the robustness of a biological system when confronted with a diverse environment, and also lack the real-time processing power of even the simplest vertebrate retina. For the relatively simple task of identifying and tracking moving objects, man-made devices fall short of the biological systems they are designed to mimic.

The goal of our research effort has been to extract and understand the engineering principles underlying natural vision systems and to apply that knowledge to designing better image processing hardware. We are focusing on the retina because research has shown that some animals possess enough image processing "wetware" to detect and track moving objects using only a thin layer of cells at the back of the eyecup (the retina).

The vertebrate retina is more than just a simple light sensor. It is a complex sensor-processor device that transforms the incoming light signal before transmitting it to the visual cortex and other subcortical regions. The retina's full range of functions are presently unknown, but it is clearly involved in dynamic range adjustments, edge enhancement, color preprocessing, and change detection. The retina has five main cell types (photoreceptors, horizontal cells, bipolar cells, amacrine cells, and ganglion cells) and two synaptic layers, the inner and outer plexiform layers, where the processes of these retinal cells interact to produce nontrival signal transformations. The outer plexiform layer handles spatial processing and dynamic range adjustments, while the inner plexiform layer is involved in change detection and temporal processing. A detailed description of the anatomy and physiology of the vertebrate retina can be found in Dowling [?]. We must emphasize that we are not trying to duplicate the biological retina. Rather we have borrowed several design principles from the retina (especially the outer plexiform layer) to solve a specific image processing problem.

Our model consists of three major components as shown in the block diagram of Figure 1: an artificial retina, augmented by a background removal network, and an image enhancement network. Processing throughout the model is performed on analog data, eliminating the need of analog-to-digital and digital-to-analog conversion.

The artificial retina is based on our previous work involving the use of a retina-like model for detecting moving objects against a fixed background [?], and consists of two parts: a photodetector array analogous to photoreceptors found in biological vision systems; and an image conditioning network that mimics the function of horizontal and bipolar cells in the biological retina. The photodetector array is a mosaic of photosensitive devices that convert light into an electrical signal. Unlike a CCD which produces discrete frames of time-averaged data, the photodetector array produces a continuous, time-varying image.

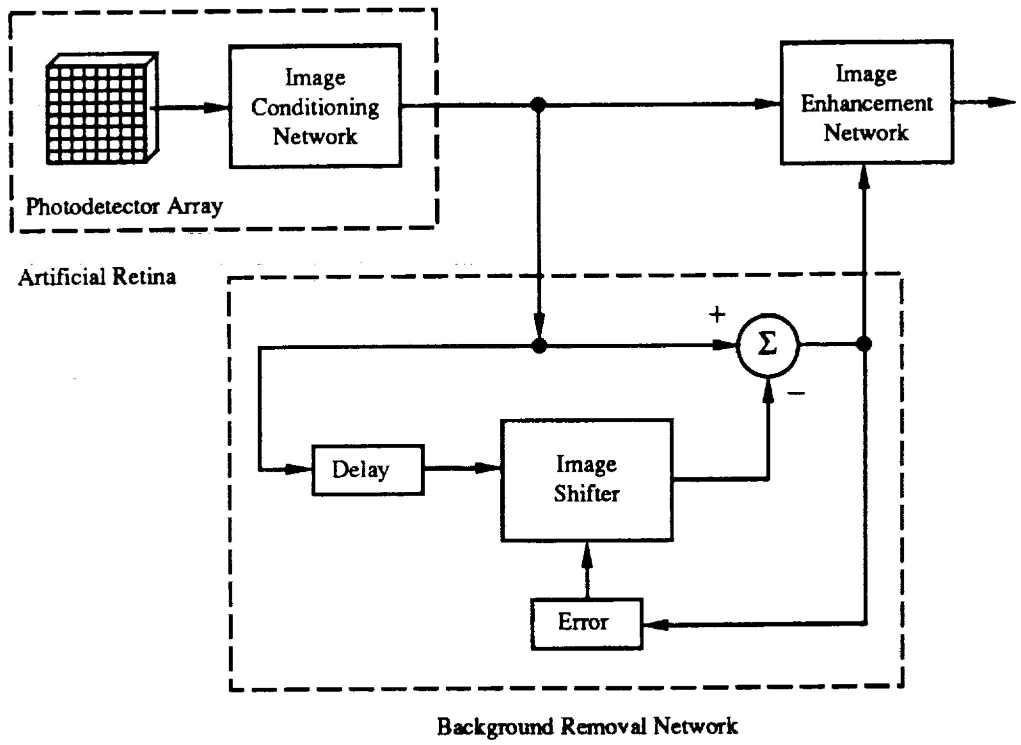


Image Conditioning Network
Photodetector Array
Artificial Retina
Image Enhancement Network
+
Σ
−
Delay
Image Shifter
Error
Background Removal Network

Figure 1: This block diagram shows the three major components of our model: an artificial retina; a background removal network; and an image enhancement network. The artificial retina converts light into a continuous, time-varying image, and conditions this image with amplification and spatial-temporal noise reduction. The background is removed by network layers which subtract a shifted, time-delayed image from the conditioned image, and the result is used to enhance the output of the artificial retina. Further analog and digital processing can be performed on the enhanced image to meet application-specific requirements.

The output of the photodetector array is amplified by the image conditioning network, which also provides temporal and spatial noise reduction.

The background is subtracted from the artificial retina output by the background removal network. To perform this operation, output from the artificial retina is first delayed by a low-pass temporal filter (depending on the application, this delay can either be fixed or variable). The delayed image is then spatially shifted by a neural network layer. The adjustable weights of this image-shifting network determine the total spatial displacement. A difference image is then formed by subtracting the delayed, shifted image from the artificial retina output. If the weights of the image-shifting network are adjusted correctly, the background in the image shifter output will be aligned with the background in the artificial retina output, and the backgrounds will cancel in the difference. Objects having a velocity difference with respect to the background will leave a negative trace at the trailing end of the object and a positive trace at the leading end.

An error feedback network modifies the weights of the image-shifting network to achieve and maintain background alignment. The drift error is determined by rectifying the difference image and summing the rectified values (1).

$$|E|^2 = \int_{Image} (I(\vec{r},t) - I(\vec{r} + \overrightarrow{offset}, t + delay))^2 \tag{1}$$

As the backgrounds are shifted towards alignment, the error decreases; as the backgrounds are shifted away from alignment, the error increases. In a one-dimensional case, the derivative of the error with respect to the spatial shift will determine the required shift direction necessary to bring the backgrounds into alignment (see Figure 2). For a two-dimensional case, the gradient of the error with respect to the X/Y spatial shift will determine the shift direction. The magnitude of the gradient scaled by a feedback gain can be used to determine the shift distance. To allow a more detailed analysis of the error feedback, we have been studying images moving in a single dimension only.

It is possible to estimate the limits on the accuracy of the offset determination due to the use of this simple, aggregate error signal. (More complex and computationally expensive shift error measures are possible, such as calculating a pixel-by-pixel brightness correlation function). To determine the effect of the background clutter, we can compute the change in the sensitivity of the error signal for different background spatial frequencies. If we assume the background is a one dimensional sinusoidal grating, then the magnitude of the error signal (as a fraction of its maximum possible value) is given by (2).

$$|E|^2 = 1 - \cos(\pi * \frac{offset\ error\ (pixels)}{background\ wavelength\ (pixels)}) \tag{2}$$

This result indicates that for very low spatial frequency backgrounds the error signal due to an offset of a single pixel will be extremely small, (e.g. 0.03% for clutter with a wavelength of 128 pixels) and will limit the accuracy of the offset optimization. However, for the applications we are studying, the background will be rich in spatial frequencies, and this error signal is quite adequate (e.g. 5.0% for clutter with a wavelength of 10 pixels).

It is important that the model be robust to environment and sensor noise. An advantage of the aggregate error signal used in our algorithm is that temporally uncorrelated noise will be averaged out in the sum over the image. To a first approximation, the variance of the error signal will be lower than the variance of the raw image signal by a factor proportional to the number of pixels. (The true statistics are somewhat more complicated since the error calculated at each pixel is rectified). Hence, random noise in the signal should cause little degradation in the optimization of the offset.

While the moving background is eliminated in the difference image, an object being tracked can take on a complex spatial structure that requires further processing prior to final detection. The purpose of the image enhancement network is to perform some of this processing on the difference image and use the processed result to enhance the retina output. The processing performed by the image enhancement network is application dependent. In our implementation, the image enhancement network performs a low-pass spatial filter on the rectified difference image and multiplies this result with the output of the artificial retina. We envision that further processing will be performed to meet application-specific requirements. For example, a readout that multiplexes and digitizes the analog image followed by digital processing, such as the Automatic Centroid Extractor (ACE) chip [?], will be necessary for a complete real-time tracking system.

Although this processing model is very versatile, there are certain limitations imposed by the general approach and the model in its current form: objects to be tracked should have a different velocity than the background; objects should typically fill only a small portion of the total field of view (FOV); and the velocity of the background must be relatively constant across the FOV. If an object has the same velocity as the background, it cannot be distinguished from the background by its motion. If an object fills a significant part of the FOV, its contribution to the total scene will bias the background motion adaptation. In the worst case, the object fills so much of the FOV that it essentially becomes the background. Finally, if the background velocity is not constant across the entire FOV, the image shift will be misaligned and portions of the background will be visible in the output. We are currently evaluating techniques to overcome these limitations.

3 Simulation Results

We wrote a simulator which allowed us to evaluate and explore variations of the retina model. To preserve the analog characteristics of the model and provide the necessary flexibility for variations, we chose to implement an abstraction of an electronic prototyping breadboard. Elements of the model are represented as analog circuit modules which can be "plugged in" and "wired" to other modules on the breadboard. A well-defined interface simplifies the task of writing new circuit modules, and existing modules can be grouped together to provide arbitrarily complex modules.

The image data used in evaluating our circuit designs was derived from a database of real and synthetic imagery. The synthetic data includes various simulated cloud scenes generated with a fractal program, earth scenes generated by a ray-tracing program, and a frequency modulated (chirped) two-dimensional sinusoid. Our real data is comprised of a sequence of images looking down upon the Earth from the Space Shuttle Challenger during the deployment of the Long Duration Exposure Facility (LDEF) satellite (Shuttle mission STS 41C). A sequence of images representing the output of the photodetector array was produced by spatially sampling a single frame of a database image (interpolation was used to obtain subpixel velocities).

One application of our software simulator in evaluating a given design is to determine the open-loop (no feedback) response to a moving background. Figure 2 illustrates the open-loop response of the design presented in this paper to a cloudy earth background moving at a velocity of -0.25 pixels per delay (this image sequence was derived from the Space Shuttle data). The error feedback to the image-shifting network was disabled, and the weights of the image shifter were initialized to values that yielded the desired spatial offset. A probe was inserted in the circuit to save the error values to a file, and the simulator was run. Next, the average error for the entire run would be computed, and another run would be performed with a different spatial offset. Performing a set of such simulation runs over a range of spatial offsets yields a curve such as that shown in Figure 2. The minima of the curve occurs at the spatial offset resulting in maximum background cancelation (since the background in the delayed image of the example is displaced by -0.25 pixels, a $+0.25$ pixel offset is required to bring the output of the image shifter in alignment with the output of the artificial retina). For spatially simple backgrounds, the open-loop error curve has a straight "V" shape. The small bends and kinks in the curve of Figure 2 are a result of the complex spatial structure of the clouds in the moving background.

We have also used the simulator to study the behavior of the closed-loop circuit. In this mode, the simulator is simply run on a data sequence, and the spatial shift of the image-shifting network is controlled by the error feedback network. The initial value of the image shifter's weights are all zero (no spatial shift), but quickly begin to change to adapt to the background motion. Similar to the open-loop study, probes are inserted into the circuit in order to save signals and images to files for post-simulation analysis.

An example of the circuit's closed-loop performance using the Space Shuttle data with a superimposed moving object (small Gaussian blob) is shown in Figure 3. The background in the circuit input (Figure 3a) moves -0.25 pixels during the delay period of the background removal network. The object, located above and to the left of the arrow in Figure 3a, moves $+0.25$ pixels during the same time interval, and has a relative intensity of fifteen percent with respect to the peak-to-peak background intensity variation. Figure 3b demonstrates the circuit's ability to remove the moving background in the input. Although the object is not easily distinguished in the input, it can clearly be seen in the output.

Stability and noise sensitivity are important concerns with systems involving feedback. Our current circuit design contains no damping elements. Appropriate choices of the feedback gain and amplifier cutoff/saturation levels avoid wild instabilities. However, after the system has adapted to the motion of the background, we observed that it tends to fluctuate slightly around the optimal spatial offset value.

In addition to the simulations described above, which were performed in batch mode,

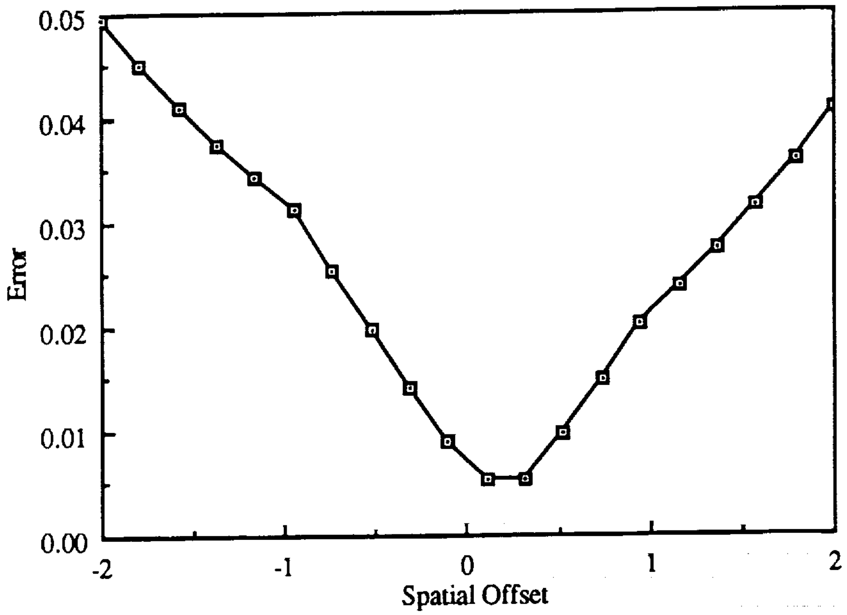
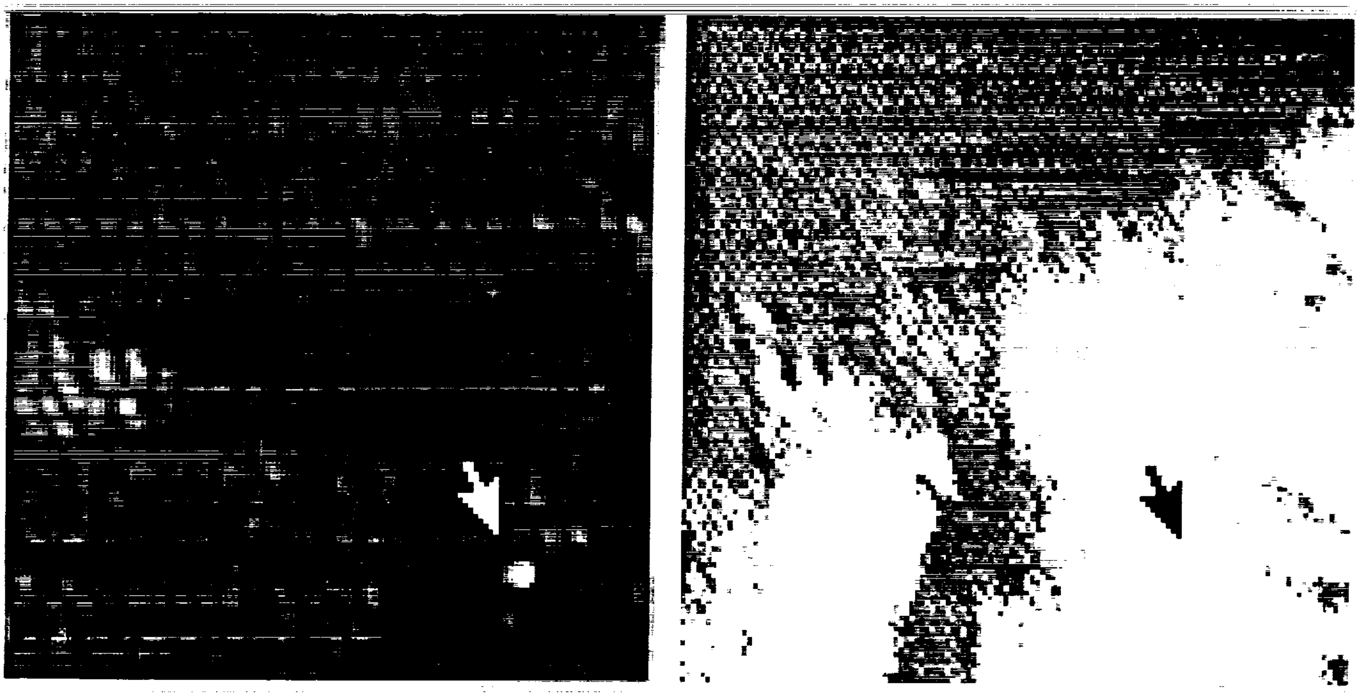


Figure 2: The system open-loop error in response to a background moving at a rate of −0.25 pixels per delay period. As the spatial offset introduced by the image-shifting network approaches the complement of the background displacement, the error between the shifted, delayed image and the unshifted image approaches a minimum. Since the background in this case has a velocity of −0.25 pixels/delay, the optimum spatial offset is +0.25 pixels. The derivative of the error is used to correct the weights in the image-shifting network to minimize the drift error.



(a) (b)

Figure 3: (a) An input image taken from a sequence of data used in computer simulation, and (b) the corresponding enhanced output image. The earth background in the input image is moving at a rate of −0.25 pixels per delay period. A superimposed object moving at a rate of +0.25 pixels per delay period is not easily distinguished in the input image, but can clearly be seen in the upper-left corner of the enhanced output image. The relative intensity of the object with respect to the peak-to-peak background intensity is fifteen percent. Initially, the spatial offset of the image-shifting network is set to zero, and after a few simulation steps, the weights of the image-shifting network adapt to the background motion.

we also implemented a real-time version of the algorithm on a PIPE image processing computer [?]. The PIPE consists of eight processors operating in parallel, each of which can perform complex operations on several frames of data in 1/60th of a second, and an ISMAP board which sums all pixels in a single image. The input to the PIPE implementation came directly from a tripod-mounted video camera. The output was displayed on a video monitor. The scalar error signal, which was computed by the ISMAP board, was passed from the PIPE to an IBM PC where the adaptation algorithm determined the new image-shifting network weights which were then passed back to the PIPE.

This real-time PIPE implementation allowed us to test the performance of the algorithm under real-world conditions (environment noise, sensor noise, and jitter caused by vibrations in the camera), and also determine the dynamic offset adjustment for a wide variety of backgrounds. As demonstrated by a video-tape we made using the PIPE, the algorithm performs well under these conditions.

4 Z-Plane Implementation

A typical optical sensor system includes optical elements, a planar array of detectors, and a CCD to multiplex the detector signals into a single signal. In such systems, processing on the focal plane is usually limited to the integration, amplification, and serial readout of the detector signals. Many operations that are currently performed off the focal plane on the digitized detector signals, such as spatial and temporal filtering, would have significantly higher performance if they could be performed in parallel directly on the continuous analog detector signals. Recent advances in fabrication and packaging technology provide the ability to stack analog or digital processing chips together and bond the stack onto the back of a detector array (the "Z" dimension) [?]. Using this technique, hardware that can exploit continuous analog image signals may now be sandwiched between the detector array and readout electronics to form a compact, cube-like image processing device.

The process of manufacturing a Z-plane module consists of thinning integrated circuit (IC) wafers to a desired thickness by precisely grinding the IC substrate, separating the IC wafer into individual circuit dies, laminating the circuit dies into a stack, forming external connections to the laminated circuits, and bonding the stack to the detector array. Z-plane modules with detector array sizes of 128x128 have been achieved, and arrays of up to 256x256 elements are in the range of current Z-plane technology. Larger focal planes have been constructed by tiling the focal plane with Z-plane modules, and a specific Z-plane implementation has been shown to have superior signal-to-noise characteristics, provide more data processing, and consume less power than a comparable CCD implementation [?].

The real-time processing required for our applications currently cannot be achieved on a conventional digital computer; however, our model maps nicely to the parallel, pipelined structure offered by Z-plane technology. Figure 4 illustrates a preliminary Z-plane packaging concept for our processing model. A photodetector array is bonded onto the first layer which implements the image conditioning network. The background removal network is

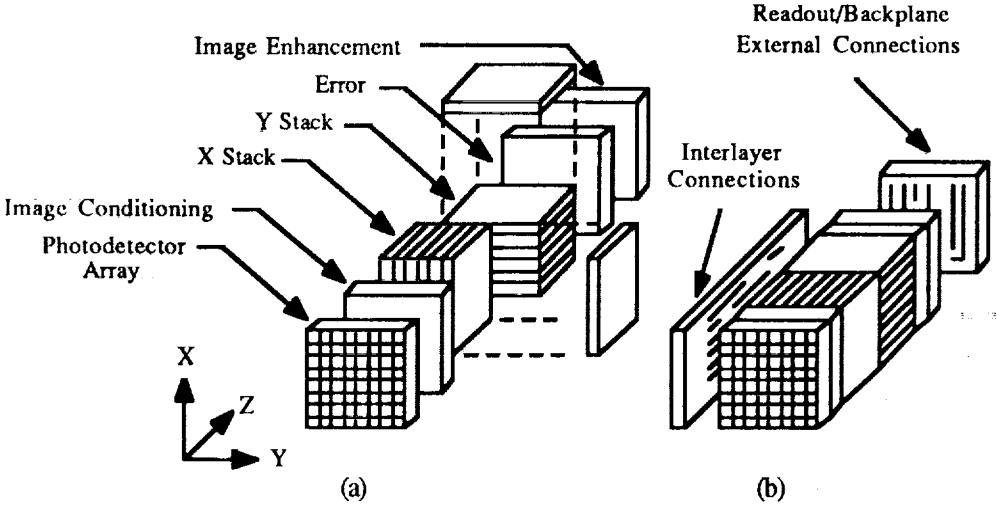


Figure 4: A preliminary Z-plane design of our model. (a) The exploded view shows the distribution of processing modules along the Z axis. The first layer is a photodetector array. An array of amplifiers and a spatial-temporal averaging network that mimics a portion of the outer retina is implemented in the following layer. The background removal network is partitioned among the next three layers (the X and Y processing stacks perform the delay and spatial shift, and the next layer computes the difference and error). Image enhancement is performed in the last processing layer. (b) A partially assembled cube illustrates the use of cube faces for interlayer and external communication.

partitioned into three layers. The first two layers implement the delay and image-shifting network. The first of these layers performs a delay and weighted sum of the inputs in the X direction, and the second layer completes the weighted sum in the Y direction (the stacking of individual IC dies for these two layers are shown in Figure 4a). The third layer of the background removal network implements the difference and error computations, and the last layer performs image enhancement. The faces of the assembled cube provides additional interlayer communication (such as the feedback signals to control the image shifter), and a readout module would be bonded to the back of the cube to provide an external interface. The multiplexed readout from such a device could be either analog or digital, depending on the nature of the readout module. Note that all signals up to the readout module are analog.

5 Summary and Future Work

We have presented an analog retina model which adapts to background motion in order to enhance objects with velocities different from that of the background. The results we have obtained from computer simulation demonstrate the model's ability to perform such enhancement for one-dimensional motion. We also presented a hypothetical implementation of our model using "Z" focal plane technology.

Since the one-dimensional simulation results are very promising, we are now proceeding to study the remaining research questions to be answered before creating a more detailed design to be implemented with Z-plane technology. We are currently investigating backgrounds with two-dimensional motion. Although in principle this is a straight forward extension to the current model, the error minimization is now in two-dimensions and may be much more sensitive to system control parameters.

Another important issue is how the system will perform in the presence of external spatial-temporal noise and internally generated noise (such as noise produced from analog component drift or component nonuniformity). Some of the noise will be removed by the artificial retina and by the implicit spatial averaging in the error signal, but we must verify that the remaining noise does not cause the optimization of the spatial offset to become unstable or experience significant drift.

We are also evaluating alternate image enhancement strategies (many of these would naturally be specific to a given application) and techniques to handle significant background velocity differences over the FOV.

6 Acknowledgements

We would like to thank the FMC Corporation for providing us access to their PIPE image processing computer. This work has been performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

References

[1] J. Carson, "Applications of Advanced "Z" Technology Focal Plane Architecture," Proc. SPIE, vol. 930, pp. 164-182, 1988.

[2] J. Dowling, *The Retina: An Approachable Part of the Brain*, Belknap Press of Harvard University Press, 1987.

[3] F. Eeckman, M. Colvin, and T. Axelrod, "A Retina-like Model for Motion Detection," Proc. IJCNN, vol. 2, pp. 247-249, 1989.

[4] T. Axelrod, T. Tassinari, G. Barnes, and K. Cameron, "A VLSI Centroid Extractor for Real-Time Target Tracking Applications," Proc. SPIE, vol. 1154, pp. 306-312,

1989.

[5] J. Martin and T. Sangston, "Z-plane Comparison with CCD for Lightning Mapper Application," Proc. SPIE, vol. 1097, pp. 16-27, 1989.

[6] Randy Luck, *An Overview of the PIPE System*, ASPEX Inc., 536 Broadway St. 10'th Floor, New York, NY.