

Dynamically Reconfigurable Vision with High Performance CMOS Active Pixel Sensors (APS)

Bedabrata Pain, Chao Sun, Chris Wrigley, and Guang Yang

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA, USA
bpain@jpl.nasa.gov

Abstract

A high-performance computational imager based on integration of CMOS active pixel imager array with on-chip signal processing and control via a new column-parallel architecture is presented. Unlike CCDs and other CMOS imagers, the sensor is capable of providing simultaneous image data from three flexible, partially overlapping, dynamically reconfigurable regions of interest (ROIs) with different spatial resolution over a large field-of-view (FOV) to achieve data-efficient operation with high update rates. The power dissipation depends upon reconfiguration modes, but is small even in the worst case. Unlike conventional computational imagers, the integration of processing circuitry does not trade with radiometric accuracy, nor does it impose any limit to scaling the imager size to larger format and smaller pixels. By eliminating the need for mechanical pointing through floating-fovea support, the dynamically reconfigurable vision sensor (DSRVS) enables a low-power staring vision system. By maximizing system throughput through intelligent data reduction and through real-time programmability, it meets the diverse and conflicting requirements of search, identify and track vision-modes, and represents an efficient platform for high-bandwidth automatic target acquisition and tracking.

Keywords

CMOS imager, computational imager, low-power VLSI

INTRODUCTION

Vision systems for surveillance, vision-based navigation, automatic target recognition (ATR), and multi-target cueing applications are presented with an overwhelming amount of visual information. The complexity arises from the fact that it concurrently carries out a number of diverse visual tasks, such as search, detection, recognition, and multi-target tracking. Search requires wide field-of-view (FOV), tracking requires fast frame rate data output from regions of interest (ROI), recognition requires high spatial resolution, while multi-target cueing requires all three of them concurrently. These cannot be easily handled by imaging systems with conventional image sensors due to the serial nature of pixel access and the enormity of the data volume. For example, a large FOV system consisting of a million pixels, operating with an update rate of 1 kHz digitized to 10 bits

will require a data output rate at a prohibitively high data rate of 10 Gbits per second, not to mention the enormity of data storage needed. On the other hand, a typical ATR scene in which the interesting information occupies only about 0.3% of the overall FOV. Thus, there is a great scope for improvement of sensor performance if the ROI can be identified in real-time and the sensor be smart enough to output data only from the region of interest, saving valuable time, and sparing costly processing and storage resources [1].

In order to solve these complex problems, biologically inspired imaging systems have traditionally attempted to incorporate local processing circuits inside the imaging pixels. Such approaches have led to the development of novel special-purpose imagers but with unacceptable image quality as a result of increased pixel size, reduced optical fill-factor, and limited imager format [2, 3]. Another approach is the use of foveal vision imagers that allow acquisition of image with varying spatial resolution that is coarser at the periphery and more refined at the center (the fovea). Although imagers with the pixel sizes scaled and organized in a foveal topology have been demonstrated [4, 5, 6], such devices provide only serial access, and consist of hard-wired (i.e. non-programmable) acuity variation. Consequently, such a vision system requires mechanical pointing, limiting its utility for miniature and high-speed operation. On the other hand, the incumbent imaging technology – charge coupled devices (CCDs) – is capable of excellent image reproduction, but is incompatible with CMOS processing, and hence of integrated sensor implementation [7]. Furthermore, CCDs are high capacitance devices, requiring multiple high voltage clocks and biases, while providing only serial output without random access, and hence are unsuitable for active vision applications.

The advent of CMOS active pixel sensor (APS) technology [8] shows great potential for overcoming both the above-mentioned limitations by integrating in one-chip both high performance imaging array (low noise, high responsivity) and embedded real-time signal processing and control circuits for optimum real-time active vision system implementation. New sensor technologies and architectures are needed for enabling such integration. In this paper, we present a real-time analog signal processing module integrated on the same chip as an imaging array to carry out recon-

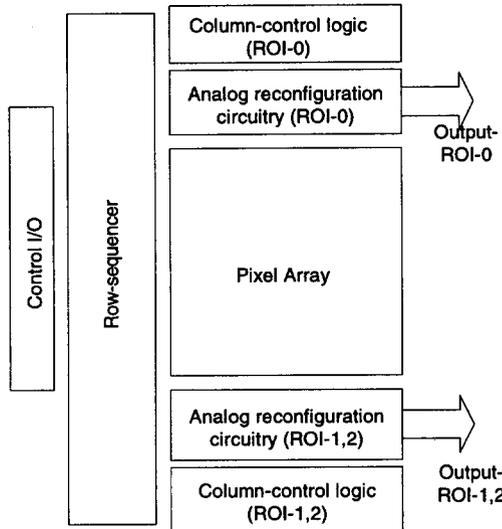


Figure 1: DRVS architecture with integrated high performance pixel array, column-parallel analog processing circuits, and column-parallel digital control circuits

figurative vision for high-speed object acquisition and tracking.

RECONFIGURABLE VISION

Dynamically Reconfigurable Vision Sensor (DRVS) architecture is shown in figure 1. It consists of a two-dimensional pixel array integrated in a column-parallel fashion to a low-power analog processing circuits that are in turn controlled by column-parallel digital control circuits. The imager is capable of multi-resolution, multi-ROI imaging at high speed. It is capable of simultaneously providing data from three partially overlapping ROIs, with the locations and resolutions of the ROIs being dynamically reconfigurable by the user. By allowing placement of the high resolution window (“floating fovea”) anywhere within the FOV and by allowing real-time programming of its resolution, the imager enables a low-power staring active vision system that meets the diverse and conflicting requirement of search, identify and track modes. Thus, the sensor can simultaneously carry out scanning over a large FOV, and tracking over a small FOV at high update rate of $> 1\text{kHz}$. Due to the use of column-parallel architecture, smart resolution adaptation is carried out without compromising imager performance, unlike that obtained in conventional biologically inspired vision.

Pixel design

In order to minimize motion artifacts, the DRVS operates in snapshot mode (i.e. simultaneous exposure for all pixels), instead of a conventional rolling-shutter mode of operation. The snapshot photodiode pixel is implemented by adding an in-pixel storage capacitance (C_p) that acts both as a frame buffer memory and as a sense node as shown in figure 2. In this case, the photodiode converts the photons to electrons for a specified amount of exposure time. Follow-

ing an exposure, the storage capacitance is reset by momentarily pulsing $\Phi_{\text{RST-C}}$. This is followed by simultaneous sharing of charges between the photodiode and respective storage capacitance by momentarily pulsing Φ_{share} . The exposure time can be varied independently of the frame read time by choosing the duration for which $\Phi_{\text{RST-D}}$ is held low.

The pixel sense nodes are readout in a row-at-a-time fashion. The pixel amplifier, which is a simple selectable source-follower with a common-column load, draws current only when a particular row of pixels is selected, and the output is sampled on a per-column basis at the bottom of the column. Differential readout of the pixel enables cancellation of pixel-to-pixel offsets MOSFET threshold voltage (V_T) mismatches, and suppression of flicker noise.

Radiometric accuracy requires that both C_p and C_d are reset to the same potential. In order to achieve this without sacrificing dynamic range, the pixel operates in flushed-reset mode. In order to enable flushed-reset during a global reset of in-pixel diodes, a

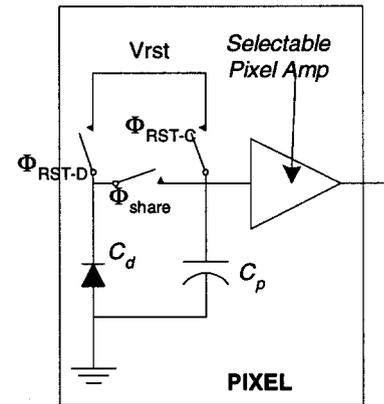


Figure 2: DRVS pixel schematic dummy pixel is selected [9].

Dynamic foveal reconfiguration

Resolution reconfiguration is carried out by averaging a block of $m \times m$ neighboring pixels to generate a super-pixel. A super-pixel represents a spatially low-pass-filtered version of the actual scene. Variable resolution can also be implemented by sub-sampling the array, which simplifies processing, but only at an unacceptable cost of image aliasing. Image aliasing leads to both false-positive and false-negative object identification.

Pixel outputs are sampled on sampled-and-hold capacitors C_{sh} in parallel during every row-phase. Averaging is carried out in a passive capacitor array [10] organized in differential column-parallel fashion as shown schematically in figure 3. The $m \times m$ block average is computed in two steps. First, the average for m pixels along the column direction for a given row of pixel values (row-averages) are computed, and stored on a set of column capacitors (C_+ and C_-). Once the computation of individual row-averages for a super-pixel block is over, average of the row-averages are computed to generate the block-average or the super-pixel value. Generation of row-averages requires each C_{sh} to be connected to its nearest neighbors by means of a switch controlled by column-generated $\Phi_{\text{row-ave}(i)}$. The size of super-pixel is determined by the bit-pattern on $\Phi_{\text{row-ave}(i)}$ that

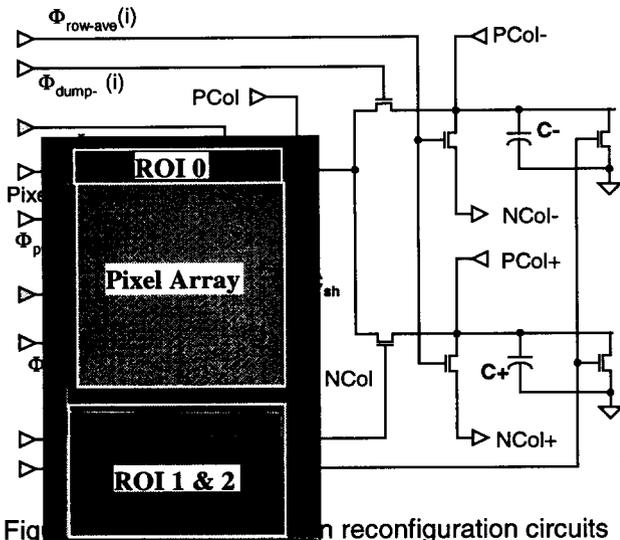


Figure 5: Photo-micrograph of DRVS chip showing reconfiguration circuits

controls the number of neighboring capacitors connected together.

Since the pixel is read out differentially, row-average is generated separately for the pixel reset values and the pixel signal values, controlled by $\Phi_{\text{dump}}(i)$'s. Block-average is generated by pulsing appropriate sets of $\Phi_{\text{col-Ave}}(i)$'s high.

For a given super-pixel size, the bit patterns for $\Phi_{\text{row-ave}}(i)$ and $\Phi_{\text{col-Ave}}(i)$ are identical, and do not vary from one row to another. The row-dump patterns [$\Phi_{\text{dump-}/+}(i)$] are shifted by one bit for every row operation. The capacitors are reset at the beginning of a block average generation by momentarily pulsing $\Phi_{\text{CR}}(i)$ high. In order to allow low-power operation, source follower bias current is turned off in columns that lie outside the user defined ROIs. This is accomplished by using the column decoded $\Phi_{\text{pwr-save}}(i)$ control signals.

Control circuits

The imager is designed to support only a selectable number of ROI resolutions, with the resolutions scaled in a binary fashion from 1x1 to 32x32. Multiple adjoining super-pixels can be grouped together to create arbitrarily large ROIs. The complexity of the column-control logic arises from the facts that the super-pixel sizes of ROIs may be different requirement, that the ROI start-addresses are arbitrary, and not integer multiples of super-pixel size, and that efficient and high-speed handling overlapping ROI addresses are necessary. The first condition implies that capacitors (C+ and C-) may be reset or shared at different instants depending upon individual super-pixel sizes.

Figure 4 schematically shows the column control block, consisting of six sub-blocks. Since separate control signals are needed for ROI-1 and ROI-2, address-range (i.e. the

extent of the window in the column direction) for each window [$W1\text{-range}(i)$ and $W2\text{-range}(i)$] is generated first from the ROI column start and end address using a nearest neighbor EXOR-based logic. The averaging bit pattern [$Ave\text{-In}(i)$] for a given super-pixel size is generated independently, and then aligned with the address-range, and the resulting bit-pattern is stored in column latches. An efficient realization of the shift-and-align operation is through the use of a 32xn diagonal switch array, where n is the total number of imager columns. The diagonal switch array uses the ROI start address to align the averaging bit-pattern with the ROI. The shifted bit pattern is then ANDed with the corresponding ROI address-range pattern to generate column-based row- and block-averaging control signals, that are subsequently latched.

Unlike the averaging control signals, the dump control signals are shifted by one bit once every row. On the other hand, no switch network for alignment of row dump address is necessary, since dumping into C+ and C- does not have to be carried out in any fixed sequence. As long as every row dump operation accesses a different set of column storage capacitors, super-pixel value will be generated correctly. The dump control signals are shifted using a recirculating shift-register, with dump control being latched separately for both ROIs after ANDing them with the respective address ranges.

On-chip power dissipation is minimized by sampling only those columns that contain active ROI data. For this purpose, an additional control line is used to disable the unnecessary column source-followers during row sampling. The pixel size is primarily limited by the need for accommodating all seven column-control lines, and is about 12 μm for advanced 0.5 μm CMOS process.

IMAGER TESTING

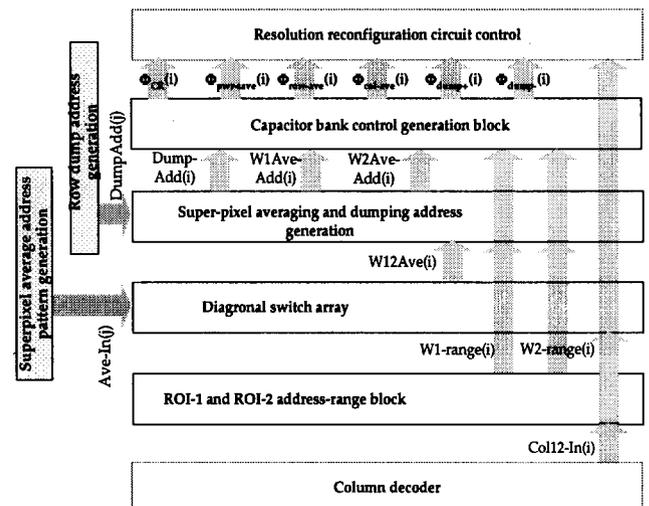


Figure 4: Schematic block diagram of column control circuit

A 256x256 snapshot photodiode-based CMOS imager in a 12 μm pitch and with on-chip variable resolution ROI generation and control circuits was implemented in 0.5 micron technology. Figure 5 shows the photomicrograph of the imager. The area occupied by the column-parallel processing and control circuits is small - being less than 3 mm in height. An important feature of the imager architecture is that the processing and control circuit size is fixed in vertical direction, and does not grow with increasing imager format, indicating its compatibility with large format imagers, unlike those implemented with biomorphic architecture [11]. Moreover, the integration of on-chip processing circuitry has not compromised any imager performance, as summarized in table 1.

Table 1. Summary of DRVS Performance

Characteristics	Value	Comments
Format	256 ²	1024 ² in design
Pixel size	12 μm	w/0.5 μm CMOS
Update rate	>1 kHz	Maximum
Quantum eff.	35%	@ 600 nm
Read noise	< 60 e ⁻	
Saturation level	> 150ke ⁻	w/3.3V operation
Fixed pattern noise	<0.1%	Norm. to saturation
Smear	<0.5%	Snapshot operation
# of windows	3	Partial overlap
Super-pixel size	1 to 32	Binary scaled
Power dissipation	5-30 mW	Worst-case

The imager shows excellent radiometric performance, and good dynamic range (~ 70 dB). The read noise is higher than that achieved with conventional photodiode pixel [12]. The increased read noise is a result of charge sharing within the diode causing charge partition noise, and separate reset of the sense node and the diode. The increase noise is, however, only ~2x compared to conventional kTC noise, partly because of proper sizing of the in-pixel capacitors, and partly because of flushed-reset, that enables sub-kTC read noise per reset [9].

Image Smear

A key issue with snapshot pixel is smear. Since the exposure time is, in general, smaller than the frame read time, it is important to prevent the photogenerated charges in the diode from diffusing into the sense node that stores the frame illumination information. Any diffusion will introduce errors in the stored value, leading to smear. The snapshot pixel contains a built-in field to prevent photoelectrons generated in the diode from reaching the sense node, thereby suppressing smear. Smear is measured as the amount of excess charge integrated on the sense node normalized by the illumination strength. Figure 6 shows a plot of smear as a function of the signal at the diode. It is clear from figure 6 that the built-in field is effective in reducing

the smear to insignificant levels (<0.5%), while previously reported results [13] for a pixel without built-in field indicate unacceptably large amount of smear (>10%).

DRV demonstration

Figure 7 shows the image of a bar-pattern captured with the DRVS chip. The full-resolution image shows excellent photometric quality. The reduced resolution images on the right demonstrate the need for signal processing (pixel averaging) during resolution reduction. Figure 7 also indicates that sub-sampling for resolution-reduction is clearly unacceptable, since it introduces false information generated through aliasing. Such problems are greatly mitigated in the super-pixel resolution reduction approach, as shown in the top-right image in figure 7. Thus, super-pixel based reconfigurable vision is of great importance for ATR and track-

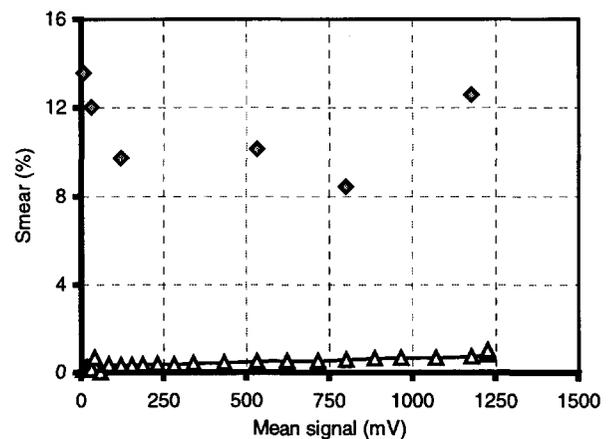


Figure 6: Measured pixel smear

ing applications.

While higher speed CMOS imagers have been developed [14], they are not suitable for ATR applications, since enormity of their data volume provided by such imagers is enough to overwhelm any downstream ATR processor. On the other hand, through flexible and dynamic imager multi-acuity reconfiguration, DRVS provides an extremely useful sensor platform for ATR applications. Figure 8 shows the image output of the DRVS operating in simultaneous wide-FOV search & scan and narrow-angle tracking mode at 100 Hz frame rate. The wide-FOV consists of a 8x8 super-pixel output from ROI-0, while the two floating variable-resolution windows output from ROI-1 and ROI-2 enable high-resolution and high speed tracking, the size of ROI-1 being 19x19, and that of ROI-2 being 32x32. While the wide-FOV data is of low-resolution, its information content is sufficient for a correlator-based processor to detect and track human faces. Thus, by simultaneously supporting flexible scanning and tracking, the chip allows maximization of system throughput by intelligent data reduction, and enables efficient and high-bandwidth ATR implementation [15].

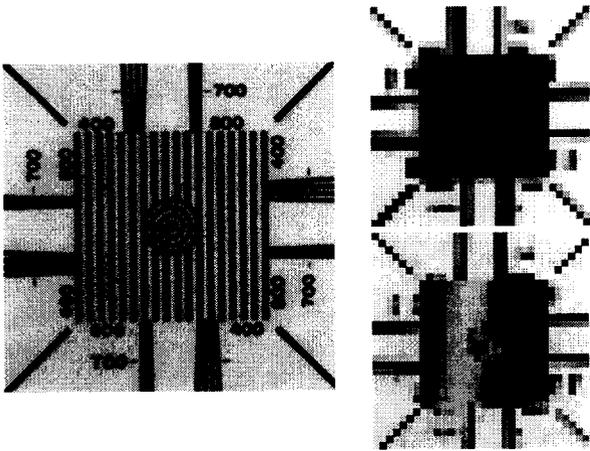


Figure 7: Demonstration variable resolution DRVS imaging. Figure on the left is captured at full-resolution, top-right with a super-pixel size of 8x8, and bottom-right with 8x8 sub-sampling

Power dissipation

Power consumption was calculated by measuring the average current flowing out of the power supply (digital and analog included), and multiplying it with the power supply voltage (3.3V). Figure 8 shows the dependence of power consumption on the pixel output rate, parametrically with super-pixel sizes. The data is captured from a full imager frame. The power consumption is low, maximum being around 30 mW. Power dissipation increases with the output rate as expected, but also with the super-pixel size. This is

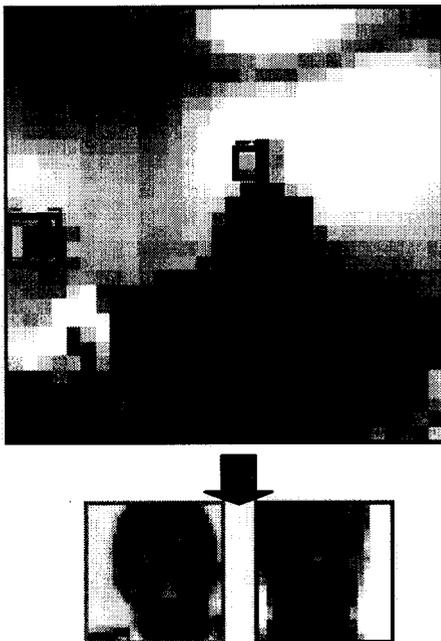


Figure 8: Simultaneous scan and track with DRVS. Image on top is from ROI-0, with two embedded high resolution windows from ROI-1 and ROI-2

because the frame rate increases for larger superpixel sizes due to reduction in the amount of data to be read out from the focal plane. As a result, for the same output data rate, power consumption increases with an increase in the super-pixel size. In fact, energy per frame reduces with increasing super-pixel size, since frame rate falls faster than the increase in power consumption per frame. At a pixel output rate of 4.16 MHz, energy per frame is $\sim 2\mu\text{J}$ for 1x1 super-pixel size, and $\sim 0.05\mu\text{J}$ for 32x32 super-pixel size.

CONCLUSION

We have presented a large-format reconfigurable multi-ROI, multi-resolution CMOS imager implemented with a new column-parallel processing and control architecture. The imager is capable of simultaneously imaging from 3 separate ROIs with user-selectable and dynamically reconfigurable resolutions. Update rates in excess of 10 kHz can be reached with less than 30 mW of power, including the processing and control electronics. A new snapshot pixel suppresses diffusion-related smear to less than 0.5%, indicating its utility for imaging with low-contrast scenes. Unlike conventional computational imagers, the incorporation of signal processing and control circuits does not de-

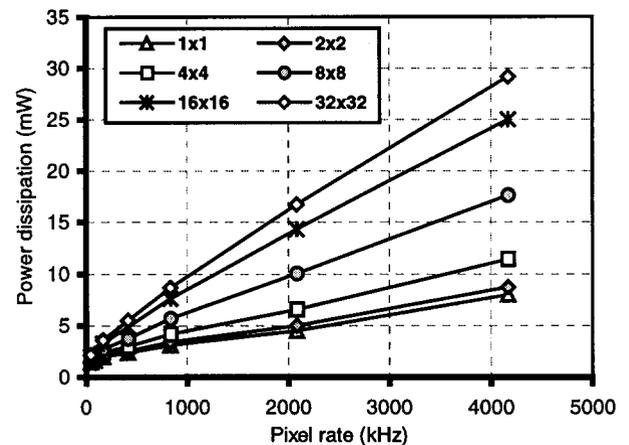


Figure 9: Power dissipation for different super-pixel sizes (full frame readout)

grade optical performance, but enables data-efficient operation with high update rates from a large field-of-view. By eliminating the need for mechanical pointing through floating-fovea support, the dynamically reconfigurable vision sensor (DSRVS) enables a low-power staring vision system. By maximizing system throughput through intelligent data reduction and through real-time programmability, it meets the diverse and conflicting requirements of search, identify and track vision-modes, and represents an efficient platform for high-bandwidth ATR and tracking.

ACKNOWLEDGMENTS

The work described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology,

and was sponsored by the Air Force Office of Scientific Research through an agreement with the National Aeronautical and Space Administration. The authors would like to thank T.J. Cunningham of JPL for invaluable technical discussion, D.J. Stack and C. Kramer of Amherst Systems, a Northrup-Grumman Company, for camera development and demonstration.

REFERENCES

- [1] P. Scott and C. Bandera, "Hierarchical Multiresolution Data Structures and Algorithms for Foveal Vision Systems," Proc. IEEE International Conference on Systems, Man, and Cybernetics, pp. 1402-1411, 1990.
- [2] M. Barbaro, P-Y. Burgi, A. Mortara, P. Nussbaum, and F. Heitger, "A 100x100 Pixel Silicon Retina for Gradient Extraction with Steering Filter Capabilities and Temporal Output," IEEE J. of Solid-State Circuits, vol. 37 (2), pp. 160-172, 2002.
- [3] A. Moini, *Vision Chips*, Kluwer Academic Publishers, Norwell, MA (1999).
- [4] R. Etienne-Cumminngs, J. Van der Spiegel, and P. Muller, "A Foveated Silicon Retina for Two-Dimensional Tracking," IEEE Trans. Circuits and Systems II, vol. 47, pp. 504-517, 2001.
- [5] G. Sandini, P. Questa, A. Mannucci, F. Ciciani, D. Scheffer, and B. Dierickx, "Image Transmission With A Retina-Like CMOS Camera," Proc. IEEE Workshop on CCD and Advanced Imager Workshop, Nagano, Japan, pp. 41-43, 1999.
- [6] F. Pardo, B. Dierickx, and D. Scheffer, "CMOS Foveated Image Sensor: Signal Scaling and Small Geometry Effects," IEEE Trans. On Electron Devices, vol. 44 (10), pp. 637-645, Oct. 1997.
- [7] A.J. Theuwissen, *Solid-State Imager with Charge-Coupled-Devices*, Kluwer Academic Publishers, Norwell, MA (1995).
- [8] S. Mendis, S. Kemeny, B. Pain, R. Gee, Q. Kim, E.R. Fossum, "CMOS Active Pixel Image Sensors for Highly Integrated Imaging Systems," IEEE J. of Solid-state Circuits, vol. 32 (2), pp. 187-198, 1997.
- [9] B. Pain, G. Yang, M. Ortiz, C. Wrigley, B. Hancock, and T.J. Cunningham, "Analysis And Enhancement Of Low-Light-Level Performance Of Photodiode Type CMOS Active Pixel Imagers Operated With Sub-Threshold Reset," in Proc. IEEE Workshop on Charge Coupled Devices and Advanced Image Sensors, pp. 140-143, 1999.
- [10] S. Kemeny, R. Panicacci, B. Pain, L. Matthies, and E.R. Fossum, "Multiresolution Image Sensor," IEEE Trans. on Circuits and Systems for Video Technology, vol. 7, no 4, pp. 575-583, 1997.
- [11] J.L. Wyatt Jr., C. Keast, M. Seidel, D. Standley, B. Horn, T. Knight, C. Sodini, H.-S. Lee, and T. Poggio, "Analog VLSI Systems for Image Acquisition and Fast Early Vision Processing," International Journal of Computer Vision, Vol. 8, pp. 217-230, 1992.
- [12] B. Pain, G. Yang, K. McCarty, J. Heynssens, T.J. Cunningham, B. Hancock, C. Sun, C. Wrigley, and C. Ho, "One Chip Digital Camera With Extended Low-Light Detection Capability," in Proc. 2000 13th. VLSI Design Conference, pp. 342-349, 2000.
- [13] G. Yang, O. Yadid-Pecht, C. Wrigley, and B. Pain, "A Snap-Shot CMOS Active Pixel Imager For Low-Noise, High-Speed Imaging," in Tech. Dig. IEEE Intl. Electron Devices Meeting, pp. 45-48, 1998.
- [14] A. Krymski, N. Bock, D. VanBlerkom, N. Tu, and E.R. Fossum, "A High-Speed, 240 Frames/Sec 4 Megapixel CMOS Imager," Proc. IEEE Workshop on Charge Coupled Devices and Advanced Image Sensors, pp. 28-31, 2001.
- [15] D. J. Stack, C. A. Kramer, T. H. McLoughlin, K. W. Sielski, G. Yang, C. Wrigley, and B. Pain, "Target Acquisition And Tracking System Based On A Real-Time Reconfigurable, Multi-Window CMOS Image Sensor," in Acquisition, Tracking, And Pointing IX, Proc. SPIE Vol. 4025, pp. 99-108, 2000.