

# A CMOS Retina based Vision System

A. Elouardi, S. Bouaziz, A. Dupret, L. Lacassagne, J.O Klein and R. Reynaud

*Institut d'Électronique Fondamentale  
Bât. 220, Université Paris Sud, 91405, Orsay Cedex, France*

Emails: [abdelhafid.elouardi@ief.u-psud.fr](mailto:abdelhafid.elouardi@ief.u-psud.fr); [bs@ief.u-psud.fr](mailto:bs@ief.u-psud.fr); [antoine.dupret@ief.u-psud.fr](mailto:antoine.dupret@ief.u-psud.fr);  
[lionel.lacassagne@ief.u-psud.fr](mailto:lionel.lacassagne@ief.u-psud.fr); [jacques-olivier.klein@ief.u-psud.fr](mailto:jacques-olivier.klein@ief.u-psud.fr); [roger.reynaud@ief.u-psud.fr](mailto:roger.reynaud@ief.u-psud.fr)

**ABSTRACT-** *This paper focuses on the VLSI compatibility of retinas, more particularly, of integrating image processing algorithms and their processors on the same sensor focal plane to provide a smart vision system on chip (SoC). The paper includes recommendations on system-level architecture and the design methodology for integrating image processing within a CMOS retina on a single chip. It highlights a compromise between versatility, parallelism, processing speed and resolution. Our solution aims to take into account the algorithms response times while reducing energy consumption so as to increase the system performances for an intelligent vehicle application.*

## 1. INTRODUCTION

Most of intelligent vehicles applications use image sensors and image processing (Muramatsu, 2002) (Bruno, 2002). These computations require a significant computing power associated to data exchange mechanisms. These functions, originally achieved by FPGA or DSP circuits, can advantageously be carried out by a microcontroller based on RISC processor coupled to an electronic Retina.

Often, images obtained from the sensors are noised because of the imperfection of the cell of capture. This induces blurriness and poor contrast of captured image. To avoid these problems, image processors are associated to the image sensors as a part of the whole vision system. Usually, two separated chips (sensing and processing) are integrated on board layout. The integration of image sensors and processing circuits on a single monolithic chip, called smart sensor, is a good solution to obtain better performance and will allow us to implement the compensation of the noised image capture for example.

Nowadays, robotics and intelligent vehicles need vision systems having fast image capture, low energy consumption, able to extract information from a visual scene so as to allow making a decision in real time. Therefore, these systems attend to be equipped with high performance hardware computing capabilities. Smart Retinas are integrated circuits in which the sensors and processing circuits co-exist (Moini, 2000). Having electronic processing elements, along with the sensor, enables to go beyond the transduction function. Most often, such circuits are for specific applications.

A silicon retina is a dedicated image sensor in which an analog and/or digital signal processing circuits are integrated in the image-sensing element (Burns, 2003) (El Gamal, 1999) (Dudek, 2000) or at the edge of the image sensor array (Ni, 2000) to

achieve some low-level image processing tasks (early-vision). Their key features are their capability to enable massively parallel computations with a rather low power.

Many approaches have been investigated: The approach presented by P. Dudek (Dudek, 2000) combines the architectural features of a general-purpose single instruction multiple data (SIMD) concept processing in the focal-plane. The implementation of an analog microprocessor in the pixel results in an Analog Processing Element (APE) and a programmable pixel-per-processor array. The suggested architecture is similar to that of retinas, but each processor has several analog memories, a communication register with neighbours pixels, and a current multiplier. Consequently, the fill factor is low and the area of a pixel remains too large ( $98 \times 98 \mu\text{m}^2$ ) to consider retinas of high resolution.

Another approach is presented by M. Arias-Estrada (Arias-Estrada, 2001). A CMOS imager is used to develop image processing architecture with the FPGA technology. The disadvantage here is the bottleneck related to the data flow between the sensor and the processing circuit and the deficiency of any kind of configuration or programmability of the array. This solution requires firm methods for a hardware implementation of the algorithm on FPGA processing circuits.

In this paper, a new processing architecture approach is presented. It highlights a compromise between versatility, parallelism, processing speed and resolution. This enables to increase the system performances.

The approach consists to set operators, usually integrated close in the pixels, at the array edge. Consequently, these functions are shared by a group of pixels, and the image processing is then carried out sequentially. This architecture results in a pixels array associated to a mixed analog-digital processors vector. Each processor is able to carry out, in situ, a wide

range of low-level image processing algorithms (Dupret, 2000, 2002). The low-level information can be then processed by a digital processor. The aim when integrating such a processor, next to image sensor in a single circuit, is to increase the fill factor and to eliminate the input output bottleneck between the sensor and the processor.

Our solution aims to take into account the algorithms response time with a significant resolution of the sensor, while reducing energy consumption for embedding reasons. The system becomes more compact and it can reach processing speeds suitable for real time applications. This paper observes the nature of image processing algorithms and categorizes them in order to find out adequate design architecture for on chip real time smart vision system.

## 2. RETINA'S ARCHITECTURE

### 2.1. Circuit Description

**PARIS** (Parallel Analog Retina-like Image Sensor) is an architecture for which the concept of retinas is modeled implementing in the same circuit an array of pixels, integrating memories, and column-level analog processors (Dupret, 2002). The proposed structure is shown in figure 1.

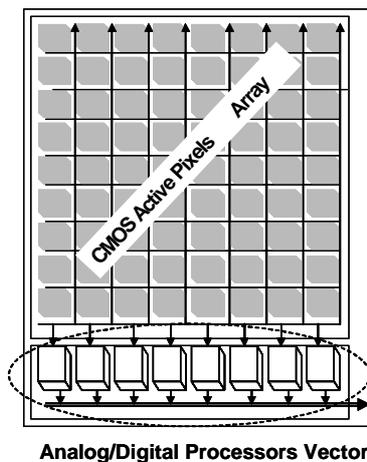


Figure 1. PARIS Architecture

This architecture allows a high degree of parallelism and a balanced compromise between communication and computations. Indeed, to reduce the area of the pixels and to increase the fill factor, the image processing is centred on a row of processors. Such approach presents the advantage to enable the design of complex processing units without decreasing the resolution. In return, because the parallelism is reduced to a row, the computations which concern more than one pixel have to be processed in a sequential way. However, if a sequential execution increases the time of processing for a given operation,

it allows a more flexible process. With this typical readout mechanism of image sensor array, the column processing offers the advantages of parallel processing that permits low frequency and thus low power consumption. Furthermore, it becomes possible to chain basic functions in an arbitrary order, as in any digital SIMD machine. The resulting low-level information extracted by the retina can be then processed by a digital microprocessor.

### 2.2. Pixels Description

The array of pixels constitutes the core of the architecture. Pixels can be randomly accessed. In some cases, the semi-parallel processing imposes to store intermediate and temporary results for every pixel in 4 MOS capacitors used as analog memories (figure 2).

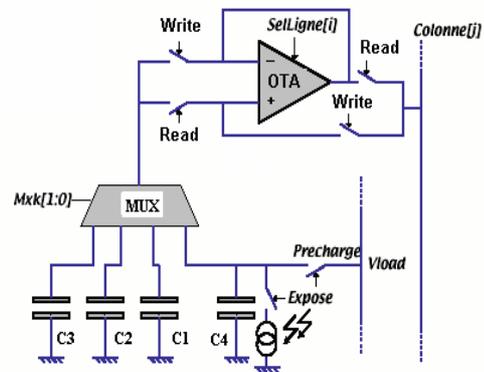


Figure 2. Pixel diagram

The selected mode for the transduction of the light is the integration mode. The photosensor is then used as a current source that discharges a capacitor previously set to a voltage  $V_{ref}$ . One of the 4 analog memories is used to store the analog voltage deriving from the sensor. Two vertical bipolar transistors, associated in parallel, constitute the photosensor. For a given surface, compared to classic photodiodes, this disposal increases the sensitivity while preserving a large bandwidth (Dupret, 1996) and a short response time can be obtained in a snapshot acquisition. The pixel area is  $50 \times 50 \mu m^2$  when the Fill Factor is 11%.

This approach eliminates the input/output bottleneck between different circuits even if there is a restriction on the implementation area, particularly for column width. Still, there is suppleness when designing the processing operators' area: the implementation of the processing is more flexible relatively to the length of the columns. Pixels of the same column exchange their data with the corresponding processing element through a Digital Analog Bus (DAB). So as to access any of its four memories, each pixel includes a bidirectional (4 to 1) multiplexer. A set of switches makes possible to select

the voltage stored in one of four capacitors. This voltage is copied out on the DAB thanks to a bi-directional amplifier. The same amplifier is used to write the same voltage on a chosen capacitor.

### 2.3. Programmable Analog Processors Vector

The pixels array is associated to a vector of processors operating in an analog/digital mixed mode. In this paper, we shall detail only the analog processing unit: APU (figure 3). Each APU implements three capacitors, one OTA (Operational Transconductance Amplifier) and a set of switches that can be controlled by a sequencer. The capacitance  $C_{out}$  plays the same role as the accumulator in a digital processor. The charge, loaded in  $C_{in1}$ , is transferred to  $C_{out}$ . According to the switches “Add” and “Sub”, the charge of  $C_{in1}$  can be added or subtracted to the charge of  $C_{out}$ . The multiplication by a constant consists in applying a voltage  $V_{in}$  to the capacitor  $C_{in1}$  while  $C_{in2}$  is reset. Next,  $C_{in1}$  and  $C_{in2}$  are connected together. Since  $C_{in1}$  and  $C_{in2}$  are at equal value, the charge in  $C_{in1}$  is divided by two. Iterating the operation  $N$  times, this step leads to a charge in  $C_{in1}$  given in the equation (1):

$$Q_{in1} = \langle C_{in1} \cdot V_{in1} \rangle / 2^N \quad (1)$$

More detailed examples of operations can be found in (Dupret, 2000).

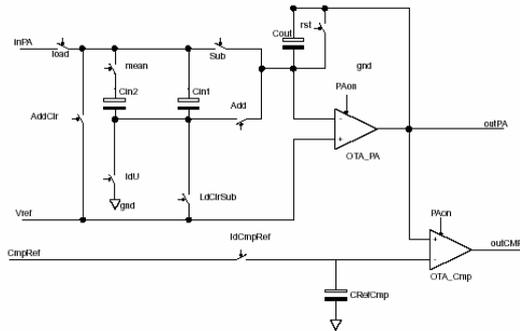


Figure 3. Analog Processor's Architecture

In order to validate this architecture, PARIS1 is the first prototype circuit that has been designed including 16x16 pixels and 16 analog processing units. This first circuit allows validating the integrated operators through some image processing algorithms like edge and movement detection. The vision chip has been design in a 0.6 $\mu$ m CMOS technology. At a first order, the accuracy of the computations depends on the dispersion of the components values. The dispersion between two APE units is 1%. Therefore, all the capacitors have the same value. A microphotography and a view of the first prototype PARIS1 circuit are

given in figure 4. The main characteristics of this vision chip are summarized in the following table:

Circuit area	10 mm <sup>2</sup>
Resolution (pixels)	16x16
Number of APUs	16
Pixel area	50x50 $\mu$ m <sup>2</sup>
Area per processing unit	50x200 $\mu$ m <sup>2</sup>
Clock frequency	10 MHz
Processing Unit power consumption	300 $\mu$ W
Pixel power consumption	100 $\mu$ W

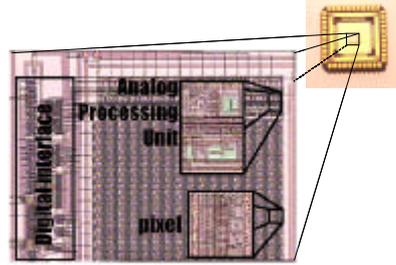


Figure 4. Microphotography of PARIS1 sensor

### 3. PARIS1 BASED VISION SYSTEM

As microcontrollers have become more prevalent and their abilities have increased, it is possible to perform pixel processing “on the fly” as the pixel values are scanned out of the retina and so a full frame buffer is not necessary. On another side, a major advantage of retinas versus a CCD camera is the ability to integrate additional circuitry on the same chip along the array of pixels. Since microcontrollers have asset of high integration, high computing power and low consumption, these characteristics make them suited for the CMOS/APS imager sensors or smart retinas (known as intelligent sensors) as a finite state machine (FSM) giving instruction to a SIMD device. Such microcontrollers support various Operating Systems and communication drivers. This suggests that it should be possible to associate a CMOS Retina with a low cost microcontroller to implement an on chip vision system.

The retina, used as a standard peripheral of the microcontroller, is dedicated to image acquisition and low-level image processing. Thanks to the analog processing units, this retina extracts the low-level information (e.g. edges detection). Hence, the system, supported by the processor, that gives the high-level information, becomes more compact and can achieve processing suitable for real time applications.

To evaluate this architecture, we have implemented a prototype based on this architecture. It is a three design parts. The first two chips are the smart retina and the microcontroller. The third part is a simple interface card implementing DAC/ADC converter (that can be integrated on the microcontroller) and decoders' circuits. The microcontroller is built around a CPU core: the 16/32-bit ARM7TDMI RISC processor. It is a low-power, general purpose microprocessor, operating

at 50 MHz, that was developed for custom integrated circuits. The aim of the evaluation is the integration of the microprocessor with the retina (PARIS1 and ARM7TDMI) on a single chip.

The advantage of this architecture remains in the parallel execution of a large number of low level operations in the array by integrating operators shared by groups of pixels (lines or columns). This allows saving expensive resources of computation, and decreasing the energy consumption. In term of computing power, this structure is more advantageous than that based on a CCD sensor associated to a microprocessor (Litwiller, 2001). Consequently, we obtain an architecture for which the PARIS1 circuit is dedicated for the regular and parallel image processing. This circuit requires a programmable sequencer, from where the advantage of integrating a microprocessor with significant capacity of computing and low fuel consumption. The control and addressing of the PARIS retina requires more ARM program computing resources to establish an FSM (Finite State Machine). PARIS retina can accept more control and addressing flow than what it is sent by the ARM programmed FSM controller. Hardware FSM version can deliver more control flow. So, our experimental results give low limit bandwidth of the retina control flow. Figure 5 shows the global architecture of the system and figure 6 gives an overview of the experimental module implemented for test and measurements.

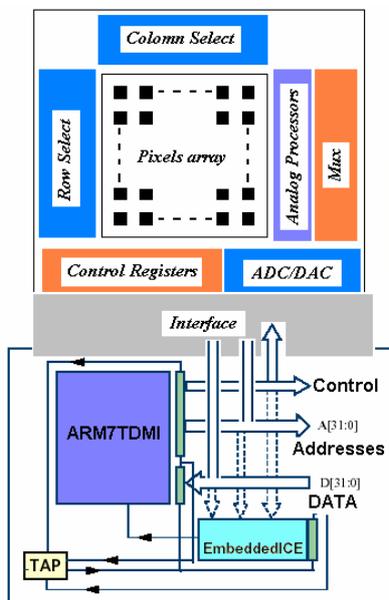


Figure 5. Global Architecture



Figure 6. PARIS1 Based Vision System

#### 4. SYSTEM EVALUATION

Parameters variations cause unavoidable non-uniformities in focal plane arrays. Since these non-uniformities change with time, calibrating sensors once is not suitable. More calibrations are then required to reduce the sensor noise and to increase the contrast for example. The primary focus of this research is to develop a single-chip imager able to provide an on-chip automatic exposure time algorithm implementing a novel self exposure time control operator. The secondary focus is to make the imager programmable, so that its performances (light intensity, dynamic range, spatial resolution, frame rate, etc.) can be customized to suit a specific machine vision application.

##### 4.1. Exposure time control

Machine vision requires an image sensor able to capture natural scenes that may have a dynamic adaptation for intensity. Reported image sensors have several constraints: large silicon area, high cost, low spatial resolution, reduced dynamic range, poor pixel sensitivity...etc. Exposure time is an important parameter to control image contrast. This is the motivation for our development of a continuous auto-calibration algorithm that can manage this state for our vision system. This avoids pixels saturation and gives an adaptive amplification of the image, which is necessary to the post-processing.

The calibration concept is based on the fact that since the photo-sensors are used in an integration mode, a constant luminosity leads to a voltage drop that varies according to the exposure time. If the luminosity is high, the exposure time must decrease, on the other hand if the luminosity is low the exposure time should increase. Hence lower is the exposure time simpler is the image processing algorithms. This will globally decrease response time and simplify algorithms.

We took several measurements with our vision system, so that we can build an automatic exposure time checking algorithm according to the scene luminosity.

Figure 7 presents the variation of the maximum grey-level according to the exposure time. For each curve, we note a linear zone and a saturation zone. Thus we deduce the gradient variation ( $\frac{d(\text{max})}{dt}$ ) according to the luminosity. The final curve can be scored out as a linear function (figure 8).

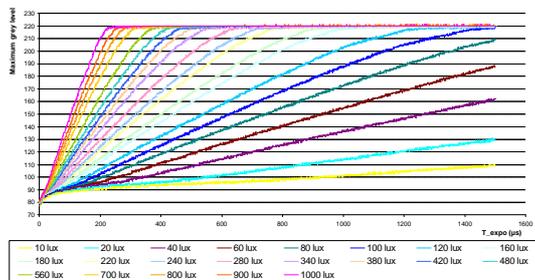


Figure 7. Measured results (Maximum grey-level versus exposure time for different values of luminosity)

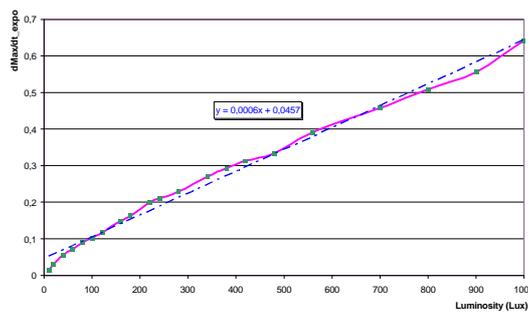
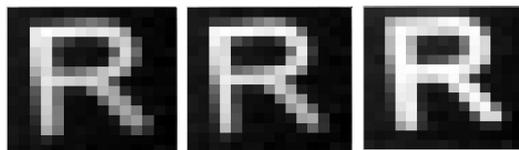


Figure 8. Gradient variation according to the luminosity

The algorithm consists in keeping the exposure time in the interval where all variations are linear and the exposure time is minimal. Control is then initialised by an exposure time belonging to this interval. When a maximum grey-level is measured, the corresponding luminosity is deduced and returns a gradient value which represents the corresponding slope of the linear function. Figure 9 gives an example of images showing the adaptation of the exposure time to the luminosity.



928  $\mu$ s, 35 Lux    512  $\mu$ s, 80 Lux    2  $\mu$ s, 1000 Lux

Figure 9. Exposure time adaptation to the luminosity

## 4.2. Image processing results

The aim of this study is to investigate what image processing algorithms can be integrated on smart sensors as a part of early vision sequences and to examine their merits and the issues that designers should consider in advance.

In this paper, we do not wish to limit implementations to application-specific tasks, but to allow for general-purpose applications such as DSP-like image processors with programmability. The idea is based on the fact that some of early level image processing in the general-purpose chips are commonly shared with many image processors, which do not require programmability on their operation.

To evaluate the functioning of PARIS architecture (each column is assigned to an analog processor), we choose an example consisting of a spatial filtering (a convolution with a matrix 3x3). The convolution kernel K used is the following:

0	-1/4	0
-1/4	1	-1/4
0	-1/4	0

Starting from an acquired image, figure 10 shows the K filtering operation result of a NxN pixels image, obtained by PARIS1, with N=16. The first line is not taken into account for the computation of the final image. Such operation is achieved in 6.8ms. The same computations require 20ms for the ARM processor. It is important to notice that the computation time increases as N for the retina and as  $N^2$  for the digital processor.



Figure 10. Original image (left) and filtered image (right)

We have successfully implemented and tested different algorithms including convolution, linear filtering, edge detection, segmentation, motion detection and estimation. Some examples are presented below. Images are processed at different values of luminosity using the exposure time self calibration.

Since the input signal is always smaller than the input range, no saturation occurs. When successive operations are performed, the coefficient applied to the input signals must be chosen so that their sum remains lower than maximum range to prevent saturations. The real limitation comes from the dynamic (the lower bound is due to noise, component mismatch, non-linearity.) of the analog processor. Finally, calibration

may be locally achieved thanks to the random access to pixels. Figure 11 gives examples of processed images.

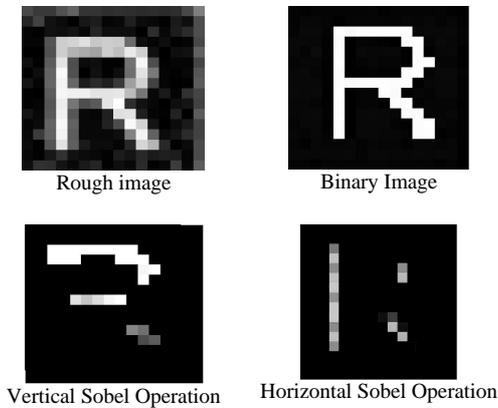


Figure 11. Examples of on chip processing images

## 5. CONCLUSION

It is concluded that on-chip image processing with retinas will offer benefits of low power consumption, fast processing frequency and parallel processing. Since each vision algorithm has its own applications and design specifications, it is difficult to predetermine optimal design architecture for every vision algorithm. However, in general, the column structures appear to be a good choice for typical image processing algorithms.

As a result, if it is possible to carry out processed images in a short time, between two processing, the relevant objects will be seen as "immobile objects". Therefore, applications involving these algorithms will be less complex and efficient to implement them on a test bench. Our implementation demonstrates the advantages of the single chip solution and contributes as a highlight. Hence, designers and researchers can have a better understanding of smart sensing for intelligent vehicles (Elouardi, 2004).

## 6. REFERENCES

- Arias-Estrada M. "A Real-time FPGA Architecture for Computer Vision", Journal of Electronic Imaging (SPIE - IS&T), Vol. 10, No. 1, January 2001, pp. 289-296.
- Bruno S., Fade: A Vehicle Detection and Tracking System Featuring Monocular Color Vision and Radar Data Fusion, Proc. of IEEE IV'2002, Versailles, France.
- Burns R., Thomas C., Thomas P., Hornsey R. "Pixel-parallel CMOS active pixel sensor for fast objects location". SPIE International Symposium on Optical Science and Technology, 3 - 8 Aug. 2003, San Diego, CA USA.
- Dudek P. "A programmable focal-plane analogue processor array" Ph.D. thesis, University of Manchester Institute of Science and Technology (UMIST), May 2000.
- Dudek P., Hicks J. "A CMOS General-Purpose Sampled-Data Analogue Microprocessor". Proc. of the 2000 IEEE International Symposium on Circuits and Systems, Geneva, Suisse.
- Dupret A., Klein, J.O. and Nshare A. "A programmable vision chip for CNN based algorithms". Proc. of CNNA 2000, Catania, Italy: IEEE 00TH8509.
- Dupret A., Klein, J.O. and Nshare A. "A DSP-like Analog Processing Unit for Smart Image Sensors", International Journal of Circuit Theory and Applications 2002. 30: p. 595-609.
- Dupret A., Belhaire E. and Rodier J.-C., "A high current large bandwidth photosensor on standard CMOS Process" presented at EuroOpto'96, AFPAEC, Berlin, 1996.
- El Gamal A. and al, "Pixel level processing Why, what and how?" SPIE Vol.3650, 1999, pp. 2-13.
- Elouardi A, Bouaziz S., Dupret A., Klein J.O, Reynaud R., "On Chip Vision System Architecture Using a CMOS Retina". Proc. of IEEE Intelligent Vehicle Symposium, IV'04. Pages 206-211. ISBN 0-7803-8311-7. June 14-17, 2004. Parma, Italy.
- Elouardi A, Bouaziz S., Reynaud R. "Evaluation of an artificial CMOS retina sensor for tracking systems". Proc. of IEEE IV'2002, Versailles, France.
- Litwiller D. "CCD vs. CMOS: Facts and Fiction". The January 2001 issue of PHOTONICS SPECTRA, Laurin Publishing Co. Inc.
- Moini A., Vision Chips. Kluwer Academic Publishers, ed. I. 0-7923-8664-7. 2000.
- Muramatsu S. and al., Image Processing Device for Automotive Vision Systems, Pro. of IV'2002, Versailles, France.
- Ni Y., Guan J.H. "A 256x256-pixel Smart CMOS Image Sensor for Line based Stereo Vision Applications", IEEE, J. of Solid State Circuits, Vol. 35 No. 7, Juillet 2000, pp. 1055-1061.