On Chip Vision System Architecture Using a CMOS Retina

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

Abstract—This paper discusses about design solution for integrating a complete vision system on a single chip. We describe the architecture and the implementation of a smart integrated retina based vision system dedicated for vehicles applications.

The retina is a circuit that combines image acquisition and analog/digital processing operators allowing the achievement of real-time image processing. Interests of vision system integration are analysed through comparisons with conventional approaches using CCD cameras and a digital processor or a CMOS sensors combined with wired algorithms on FPGA technology. Our solution will take advantages of both solutions.

I. INTRODUCTION

Most of intelligent vehicles applications use image sensors and image processing [1][2]. 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 [3]. 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 [5][6][9] or at the edge of the image sensor array [7] to achieve some simple 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. Dudeck [8] 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 (98x98µm²) to consider retinas of high resolution.

Another approach is presented by M. Arias-Estrada [10]. 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 [11][12]. 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.

II. SMART RETINA DESCRIPTION

PARIS (Programmable Analog Retina-Like Image Sensor) is an architecture for which the concept of retinas is respected by integrating in the same circuit the acquisition photo-sensors and some processing operators [13]. It presents a high degree of parallelism and a compromise between the communications and computation.

![PARIS Architecture](image1)

This architecture, shown in figure 1, is designed to support up to 256x256 pixels. We detect three key blocks: an array of memories with photo-sensors, an analog processing vector and a microprocessor. PARIS.1 is a 16x16 pixels VLSI prototype with 16 analog processors. This first circuit allows validating the integrated operators through some image processing algorithms like edge and movement detection. The analog processors execute instructions fed by a microprocessor achieving a program. Pixels can be randomly accessed. Each pixel consists of a photo-sensor and four analog capacitors acting as memories (figure 2).

![Pixel Scheme](image2)

The pixel array is associated to a vector of processors operating in an analog/digital mixed mode. The column structure of PARIS architecture implements a mixed analog/digital unit: analog processor (AP), Boolean unit (BU) and analog registers (figure 3).

![Analog-Digital Processor Unit](image3)
The analog processors execute instructions fed by a microprocessor achieving a program. The figure 4 details architecture of the analog processor. The comparator allows analog results to be converted to logical predicate signal. The analog processing unit (AP) is based on switched capacitors integrators. It contains three capacitors, one OTA (Operational Transconductance Amplifier) and a set of switches controlled by a digital processor. The capacitor Cout is used as an accumulator. Thanks to the OTA, the charge stored in capacitor Cin1 is transferred towards Cout so that it adds or subtracts.

![Fig. 4. Analog Processor Architecture](image)

Furthermore when the capacitor Cin1 is first shorted, and then put in parallel with Cin2, the charges balance among Cin1 and Cin2. As a result, the sequence of operations achieves the division by two of the present charges on the two capacitors Cin1 and Cin2. Hence, the transferred charges can be divided by two.

Pixels of the same line are simultaneously processed by the Analog Processors (AP) vector and the computing is iterated on image rows. The arithmetic operations (division, addition) are carried out in analog. The accumulation of the intermediate results is achieved in the analog processor by using the internal analog registers.

The analog processor unit integrates a comparator, so it possible to implement analog to digital conversion by software sequenced instructions. Such a structure has the advantage of an analog ALU which avoid data conversion stage. However, in spite of the analog data processing, the control data flow is digital.

### III. HARDWARE ARCHITECTURE AND IMPLEMENTATION

A major advantage of retinas versus CCD cameras technology is the ability to integrate additional circuitry on the same chip along the array of pixels.

On another side, as microcontrollers have become more prevalent and their abilities have increased, it is possible to perform simple pixel processing “on the fly” as the pixel values are scanned out of the retina and so a full frame buffer is not necessary. Since microcontrollers have asset of high integration, high computing power and low consumption, these characteristics make them required by the CMOS/APS imager sensors or Smart Retinas (known as intelligent sensors). Such microcontrollers support various Operating Systems and communication drivers. This suggests that it should be possible to associate a CMOS retina chip with a low cost microcontroller to implement an on chip vision system.

To evaluate this architecture, we have implemented a prototype based on this idea. 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 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 use in specific applications and custom integrat

![Fig. 5. Global Architecture](image)
level information (e.g., edges detection). Hence, the system, supported by the processor, becomes more compact and can achieve processing suitable for real-time applications.

The advantage of this architecture type remains in the parallel execution of a consequent number of low-level operations in the array by integrating operators shared by groups of pixels. This allows saving expensive resources of computation and decreasing the energy consumption. In terms of computing power, this structure is more advantageous than that based on a CCD sensor associated to a microprocessor. Figure 5 shows the global architecture of the system and figure 6 gives an overview of the implemented module.

IV. APPLICATIONS AND RESULTS

A. Auto-Calibration Algorithm

Luminosity variations cause unavoidable non-uniformities in focal plane arrays and other integrated sensors. Since these non-uniformities change with time, calibrating sensors once is not suitable and frequent calibrations are then required. Image calibration enables to reduce the sensor noise and increase contrast.

Machine vision requires an image sensor able to capture natural scenes that may have a dynamic adaptation for intensity. Reported wide image sensors suffer from some or all of the following problems: large silicon area, high cost, low spatial resolution, small dynamic range, poor pixel sensitivity, etc.

The primary focus of this research is to develop a single-chip imager for machine vision applications which addresses these problems, but is still able to provide an on-chip automatic exposure time algorithm by implementing a novel self-exposure time control operator. The secondary focus of the research is to make the imager programmable, so that its performance (light intensity, dynamic range, spatial resolution, frame rate, etc.) can be customized to suit a particular machine vision application.

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 globally will 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 (Δmax/Δt) according to the luminosity. The final curve can be scored out as a linear function (figure 8).

The algorithm consists to keep 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.
B. Image processing results

The basis of the smart on chip vision system concept is that analog VLSI systems with low precision are sufficient for implementing many low-level vision (image processing) algorithms, often for application-specific tasks. Conventionally, smart sensors are not general-purpose devices, but everything in a smart sensor is specifically designed for the targeted application.

General image processing consists of several image analysis processing steps: image acquisition, pre-processing, segmentation, representation or description, recognition and interpretation.

The order of this image analysis can vary for different applications, and stages of the processes can be omitted. In image processing, the image acquisition is used to capture raw images from its input scene, through the use of video camera, scanners and, in the case of smart retinas, the solid-state arrays.

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.

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.

Indeed, the aim of our work is to explore the potential benefits of using a retina plus a processor instead of the classical approach (e.g. imager associated to a DSP). We hence focus our works on the point where the gain of our approach can be significant in the low level image processing step of the image vision. Since the retina we use combines a fully random access to the pixel along with the ability to perform parallel computations, the local low level image processing tasks can be performed at high speed while the ARM processor handle the high level processing tasks.

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

$$
\begin{array}{ccc}
0 & -1/4 & 0 \\
-1/4 & 1 & -1/4 \\
0 & -1/4 & 0 \\
\end{array}
$$

Starting from an acquired image, the figure 9 shows the $K$ filtering operation result of an $N\times N$ 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.

Opposite to integration that is analogous to averaging or smoothing, differentiation can be predictable to sharpen an image leaving only boundary lines and edges of the objects. This is an extreme case of high pass filters. The most common methods of differentiation in image processing applications are first difference, gradient and laplacian operator.

The difference filter is the simplest form of the differentiation with subtracting adjacent pixels from the centred pixel in different directions. The gradient filters represent the gradients of the neighbouring pixels (image differentiation) in forms of matrices. Such gradient approaches and their mask implementations are represented with various methods: Roberts, Prewitt, Sobel, Kirsch and Robinson.

With many different local operations in image processing algorithms, these operations can be categorized...
into three major groups: smoothing filters, sharpening filters and edge detection filters.

We have successfully implemented and tested a number of 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 local calibration may be locally achieved, thanks to the random access to the pixel.

Figure 10 gives an example of images showing the adaptation of the exposure time to the luminosity and figure 11 gives examples of processed images.

<table>
<thead>
<tr>
<th>Luminosity</th>
<th>Exposure Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>35 Lux</td>
<td>928 µs</td>
</tr>
<tr>
<td>80 Lux</td>
<td>512 µs</td>
</tr>
<tr>
<td>1000 Lux</td>
<td>2 µs</td>
</tr>
</tbody>
</table>

Fig. 10. Exposure time adaptation to the luminosity

Rough image

Binary Image

Horizontal Sobel filtered image

Vertical Sobel filtered image

Fig. 11. Examples of processed images

V. CONCLUSION

The require for real time image processing applications for portable and battery-operated devices has grown swiftly in recent years, and it has motivated the research on processing architectures that support focal plane data.

It is concluded that on-chip image processing with retinas will offer benefits of low manufacturing cost, 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 the best choice for typical image processing algorithms.

We have presented the PARIS architecture and the implementation of a prototype based vision system. The goal is the integration of a processor in the retina, to manage the system and optimize the hardware resources use. We propose evaluating such a system using retinas with high resolution starting from a complex application for autonomous collision avoidance and objects tracking at real-time (figure 12): an embedded system for autonomous collision avoidance and objects tracking.

Fig. 12. Embedded system for vehicle applications

REFERENCES


