3D multi-layer vision architecture for surveillance and reconnaissance applications

Peter Földesy[+], Ricardo Carmona-Galan*, Ákos Zarándy[+], Csaba Rekeczky[++], Angel Rodríguez-Vázquez**, Tamás Roska[+]

[+]Computer and Automation research Institute of the Hungarian Academy of Sciencies (MTA-SZTAKI), Budapest, Hungary foldesy@sztaki.hu

* Instituto de Microelectrónica de Sevilla, (IMSE-CNM), Sevilla, Spain rcarmona@imse.cnm.es

[++]Eutecus, Inc, Berkeley, CA, USA. rcsaba@eutecus.com

**AnaFocus, Seville, Spain Angel@anafocus.com

***Abstract*— The architecture and the design details of a multi-layer combined mixed-signal and digital sensor-processor array chip is shown. The processor layers are fabricated with 3D integration technology, and the sensor layer is integrated via bump bonding technology. The chip is constructed of a 320x240 sensor array layer, closely coupled with a 160x120 mixed-signal processor array layer, a digital frame buffer layer, and an 8x8 digital fovea processor array layer. The chip is designed to solve image registration and feature extraction above 1000FPS.**

I. INTRODUCTION

This paper introduces the architecture a novel programmable, application specific vision system which is designed for autonomous visual navigation applications, including exploration, surveillance, target tracking. Since the target carriers are small mobile platforms (UAVs or ground vehicles) ultra compact system size and low power consumption are of crucial importance. To be able to achieve these requirements, we have selected to implement the vision system by using an advanced vertically integration technology which applies high density 3D through silicon vias (TSV) with 5 micron pitch.

Three components with enhanced functionality are combined in this 3D monolithic silicon chip (called "VISCUBE"):

- a high sensitivity sensor array;
- a programmable, fully parallel, mixed-signal topographic processor array,
- a digital frame buffer, and
- a programmable, fully parallel, digital multi-core processor array with local memories in each node.

The specialty of the VISCUBE architecture is that it combines a sensor layer and two closely coupled topographic processor array layers operating in different signal domains on a single chip. The roles of the two processor arrays are different also. The front-end mixed signal layer identifies the points/regions of interests, while the lower resolution digital processor array performs high precision foveal processing in the identified locations.

The paper describes the architecture of the VISCUBE (Section 2), the physical implementation (Section 3), and the algorithmic framework (Section 4).

II. ARCHITECTURE

VISCUBE is a focal-plane sensor-processor vision chip. It is constructed of a sensor layer, a mixed-signal fine-grain processor architecture (specialized version of Q-Eye [1]), a buffer memory layer, and a coarse-grain digital processor array (advanced version of Xenon [2]). Fig. 1 shows the high-level block diagram of the architecture and its implementation in a multi-layer silicon chip.

A. *Sensor layer*

The sensor array is built on an extra semiconductor layer, connected to the processor array using bump bonding interface. The choice of the sensor material defines the wavelength. This can range from visual range to NIR, or even SWIR. The extra semiconductor layer enables reaching close to 100 percent fill factor without eating up the top silicon tear of the 3 tier processor system.

The resolution of the sensor array is 320x240. The sensor layer contains the photodiodes only. The integration type sensor interface is implemented on the mixed signal layer, which makes relatively straightforward the usage of different sensor materials.

B. *Mixed-signal processor*

The upper silicon layer contains a 160x120 resolution programmable smart image processor cell array and sensor interface. The mixed-signal processors (or cells) serve the image acquisition, diffusion, subtraction and local maximum location map generation. In details, each cell comprises electrical interface for 4 sensors, pixel memories, linear analog diffusion unit, and local extremum detector. Each cell is

978-1-4244-3896-9/09/$25.00 ©2009 IEEE

interconnected with its 8 neighboring cells, allowing for programmable real-time image acquisition and spatial processing operations. Each processor is interfaced with four optical sensors topographically on a pitch-matched way. This means that each pixel can both sense the corresponding spatial sample of the image and process this data in close interaction and cooperation with other pixels.

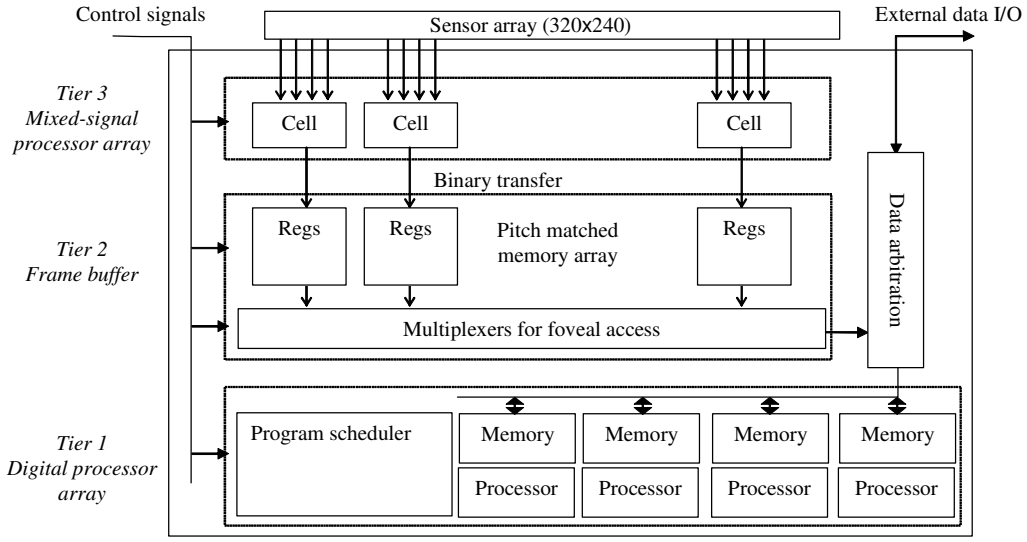

Fig. 1. Functional architecture of the VISCUBE system.

The mixed signal processor array is designed to handle both 160x120 and 320x240 and process 60x120 sized images. The spatial operators can process low resolution images. In those cases, when 160x120 sized image is processed the 160x120 image is generated by subsampling or binning the 320x240 image. The pitch of the mixed signal processor layer is 60 micron, applying 0.15 micron technology.

1) Image acquisition

The image acquisition is the integration of the photocurrent from the external photodiodes. The integration time is controllable in a wide range (from sub-microseconds to hundred milliseconds). The sensor interface applies a transconductance amplifier. The photodiodes are kept on a constant bias voltage during the integration. This bias voltage can be set externally to support various sensor materials.

2) Diffusion operator

The diffusion operator is defined on the processor grid resolution (160x120). The amount of diffusion time is controllable based on external precision clock signals. As the resistive grid based diffusion it has an:

$$e^{-\frac{|d|}{\sigma}}$$

type kernel. (d is the distance between cells) The diffusion time has a granularity of sigma 0.5.

The diffusion process can be restarted at any meaningful times, which enables further smoothed image generation and conversion.

3) Difference generation

The difference generation as the input of extremum location detection is performed on the image with different diffusion kernels. In practice, the first argument of the subtraction is a snapshot of the image during diffusion, the second argument is the final result of the diffusion.

The minimum time resolution (i.e. diffusion kernel) of the snapshots is the same the minimal diffusion time. The snapshot also can be at zero time that means original image.

4) Extremum location detection

The extremum location detection is also defined on the cell resolution grid. The output of the operator is a binary image, which is black in those pixels, which contains local maximum/minimum in a 3x3 neighborhood. In order to suppress irrelevant local extremums, a pixel is defined as maximum/minimum if it exceeds a threshold limit. This threshold value is common for all cells.

5) AD conversion

Each of the cells contains a single slope AD converter. The comparator of the AD is located on Tier 3, in the mixed signal layer, while the digital memory is located in the frame buffer layer. This requires a pitch matched frame buffer layer design with one TSV/cell interconnection.

C. Frame buffer layer

The frame buffer has three roles. It is used as the (i) digital registers of the single slope AD converters, which can be found in each cells. It is also used as a (ii) storage and communication interface between the mixed-signal and the digital layer. Moreover, (iii) it supports random access to scaled or not scaled windows of the captured and preprocessed images, which is critical in the multi-scale and fovea processing approaches.

The frame buffer layer is constructed of an array of 160x120 memory units (Fig. 2). Each memory unit is

corresponding to the mixed-signal processor geometrically located exactly above it. (The two arrays are designed in a pitch matched way.)

Each memory unit contains 6 bytes of memory and 2 bits. 4 bytes is needed to store a 320x240 sized image, because in this case, each unit/cell handles 4 pixels. The remaining 2 bytes can be used to store multiple downscaled images. The single bits are used to store the outputs of the extremum filter.

The registers of the frame buffer are constructed of twoport memories. They are written when an AD conversion or extremum operation is executed. They are read out by the digital processor array layer through a multiplexer. The multiplexer supports automatic windowing and downscaling functions to minimize IO time.

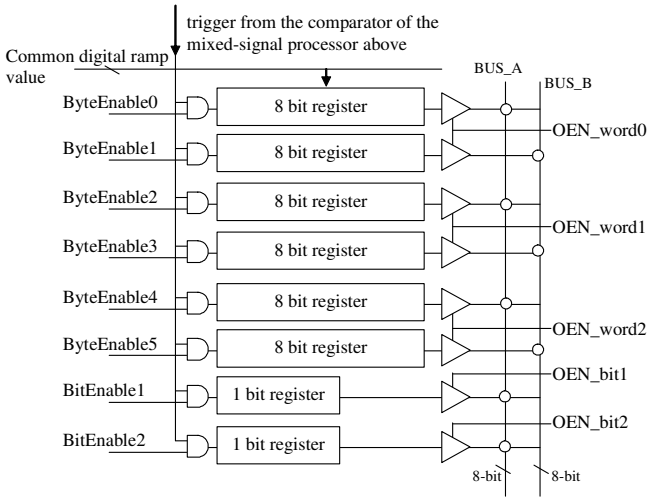

Fig. 2. Memory unit of the frame buffer layer

D. Digital processor layer

The digital processor array is intended to be used for both area of interest (window) and full frame processing as well. This 8x8 processor array is an advanced version of the Xenon [2] architecture. The distinguishing feature of this new implementation is the increased memory size (2kbytes/processor) and higher flexibility in the processed window size.

The processors can work in two different modes. In the first mode, the 64 processors are joining forces, and process a large or medium sized image (320x240, 160x120, 80x60, or 64x64) by topographically distributing the image data to the processors. The second mode is used, when we have to execute the same operation (e.g. displacement calculation of a feature point) on a large number of smaller windows (24x24 or 16x16). In this case each of the processors process one window individually. In the formal mode, the neighboring processors are exchanging data intensively, while in the latter mode, the processors are uncoupled. In both modes the processor array operates in SIMD mode.

The basic constructing element of our digital processor array architecture is the cell. The cells are locally interconnected. The processors in each cell can read the memory of their direct neighbors. There are boundary cells, which are relying data to handle different boundary conditions.

Each cell contains an arithmetic processor unit, a morphologic processor unit, data memory, internal and external communication unit. The arithmetic unit contains an 8 bit multiple-add core with a 24 bit accumulator, and 8 pieces of 8 bit registers. This makes possible to perform either 8,16, or 24 bit precision general calculations. The arithmetic unit can calculate multiplication, multiple-add, addition, subtraction, and comparison operation. Image processing primitives, like block matching, convolution, look-up table, diffusion, thresholding, rank order filters, contour detection, Sobel operator, median, etc. can be efficiently implemented by using the instruction set of the processor.

The morphology unit supports the processing of black-and-white images. It contains 8 pieces of single bit morphology processor, for parallel calculation of local or spatial logic operations, like erosion, dilation, opening, closing, hit and miss operations, etc. The detailed description of the processor array and the cells can be found in [2].

III. PHYSICAL IMPLEMENTATION

The vision system is implemented using 3D vertical integration technology. It is constructed of three silicon processor/memory layers and a bump bonded top photodiode array layer. The 320x240 sensor array layer is composed of a photodiode array structured on Silicon or maybe on other semiconductor. Bump bonding integration technology enables fill factor close to 100%, a necessary prerequisite to build truly highly sensitive sensors. Based on the final selection of the sensor materials, the VISCUBE will be sensitive in either the visible (VIS) or in short wave infrared (SWIR) range.

Below the sensor layer, there are three SOI layers (0.15 micron technology) with dense (5 micron) vertical interconnections with 3 metal layers on each. The implementation of the described architecture is currently an ongoing project. The design will be submitted in Q3 2009.

IV. TARGET ALGORITHMS

The primary application of the VISCUBE is airborne visual navigation and reconnaissance. These applications are based on segmentation. However, segmentation of an image on a moving platform requires image registration too. In this context, image registration means to find and calculate the affine transformation compensating the ego motion of the camera. Hence, the target algorithms of the VISCUBE is based on image registration as it is shown in Fig. 4. First, the registration requires the identification of the characteristic points of the image. This is done by the finding the local extremes in different scale-space dimensions Fig. 4. These points are identified by the mixed-signal layer. The feature points are sought on 160x120 resolutions. There are two reasons using reduced resolution. First of all, full resolution (320x240) mixed signal layer does not fit to the given silicon space. Moreover, according to our simulation results, the inaccurate identification of the location of the feature points does not cause any problem, because these are used to identify

the location of windows where displacement is calculated in full resolution in the subsequent steps.

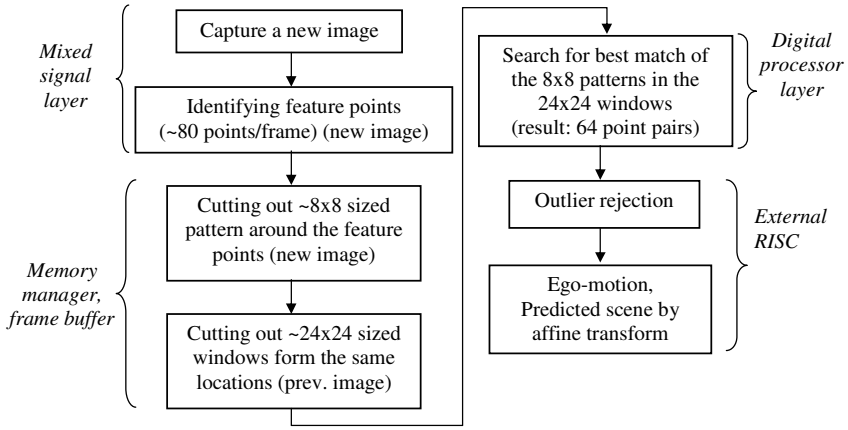

Fig. 3. Flow-chart of the image registration algorithm

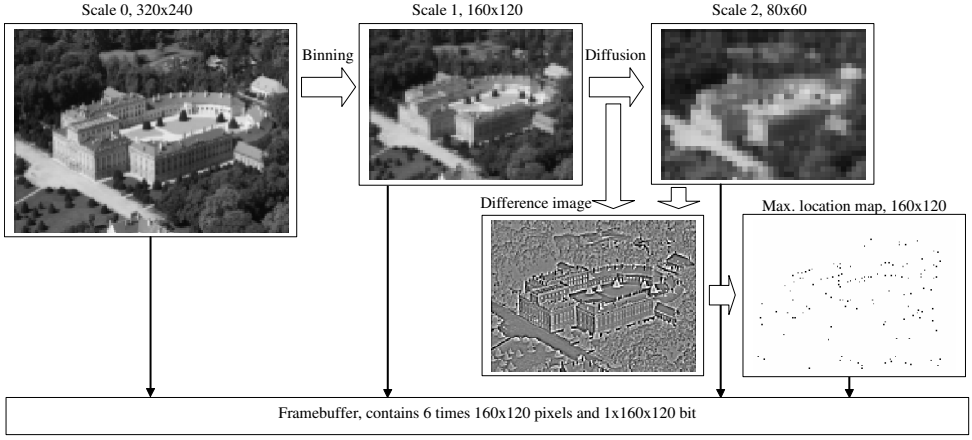

Fig. 4. The major steps of the feature point extraction

After the feature points are identified, the full resolution original image – stored on the mixed signal layer in analog form – is converted to the digital domain using the distributed AD converters, and saved on the frame buffer layer. From this location the foveal readout of the image in different scales is fast thanks to the implemented multiplexer layer.

In the next step, displacement is calculated between the actual frame and the previous frame in the locations identified by the feature points. This is done on windows cut out from the high resolution (320x240) images. The matching calculation is implemented on the digital processor layer on a way that each processor calculates displacement on one window. In the simulations, we calculated the best matching of an 8x8 pattern in a 24x24 window. The calculation time is in the order of 0.5 ms. Since all the processor is used, 64 displacements are calculated parallel during this time.

Last steps of the algorithm are the outlier rejection and the completion of the affine transform. These are done by a RISC processor, which applied as a companion processor of the VISCUBE system.

V. CONCLUSION

A new vision sensor-processor chip is introduced. The special feature of the new architecture is that it combines multiple cellular processor arrays operating in different signal domain, and processing the image in different scale and resolution. The new design is implemented by applying the latest 3D silicon integration technology.

VI. ACKNOWLEDGEMENT

The work is supported by the Eutecus ONR-BAA Co. Num N00173-08-C-4005 VISCUBE project.

VII. REFERENCES

[1] www.anafocus.com

[2] [13] P. Földesy, Á. Zarándy, Cs. Rekeczky, and T. Roska „Configurable 3D integrated focal-plane sensor-processor array architecture”, Int. J. Circuit Theory and Applications (CTA), pp: 573-588, 2008.

[3] Eye-RIS v1.2 IPL Reference Eye-RIS™ v1.2 Vision System Evaluation Kit, AnaFocu