

Digital processor array implementation aspects of a 3D multi-layer vision architecture

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

⁺Computer and Automation research Institute of the Hungarian Academy of Sciences (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. rscaba@eutecus.com

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

Abstract—Technological aspects of the 3D integration of a multi-layer combined mixed-signal and digital sensor-processor array chip is described. The 3D integration raises the question of signal routing, power distribution, and heat dissipation, which aspects are considered systematically in the digital processor array layer as part of the multi layer structure. We have developed a linear programming based evaluation system to identify the proper architecture and its parameters.

Keywords: 3D integration, UAV navigation, sensor-processor

I. INTRODUCTION

This paper describes implementation considerations of a programmable, application specific vision system which is designed for autonomous visual navigation applications, including exploration, surveillance, target tracking [2]. Since the target carriers are small mobile platforms (UAVs or ground vehicles) ultra compact system size and low power consumption are of crucial importance. We have selected to implement the vision system (called “VISCUBE”) by using an advanced monolithic vertically integration technology of 3D through silicon vias (TSV) with 5 micron pitch, and three 0.15 um feature sized SOI CMOS tiers.

The paper briefly describes the digital multi-core processor architecture of the VISCUBE (Section 2), the performance requirement of the processor layer (Section 3), and the signal and power/ground distribution (Section 4).

II. ARCHITECTURE

The digital processor array is intended to be used for both area of interest/fovea (window) and full frame processing as well. This 8x8 processor array is an advanced version of the Xenon [1] architecture. The distinguishing feature of this new derivation is the increased and uneven memory size throughout the array (0.5-2 kbytes/processor) and further memory accessing modes to facilitate the fovea processing.

The application field of the VISCUBE is airborne visual navigation and reconnaissance. These moving platform applications typically solve the registration of consequent frames. In our context, image registration means to find and calculate the affine transformation compensating the ego motion of the camera. In order to reduce computational load, our approach relies on characteristic or feature point tracking

instead of full frame registration. Furthermore, hierarchically scaled images are used during feature point identification and tracking. Further details are provided in [1].

Three components are combined in the VISCUBE to spread the computation load (Fig. 1):

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

The visual input comes from the fourth layer, bonded pixel-wise on top of the chip. The control and the synchronization of the VISCUBE will be provided by external host processor. This processor will execute the main program, initialize subroutines on individual layers, and synchronize the data communication among the three processing units of the system.

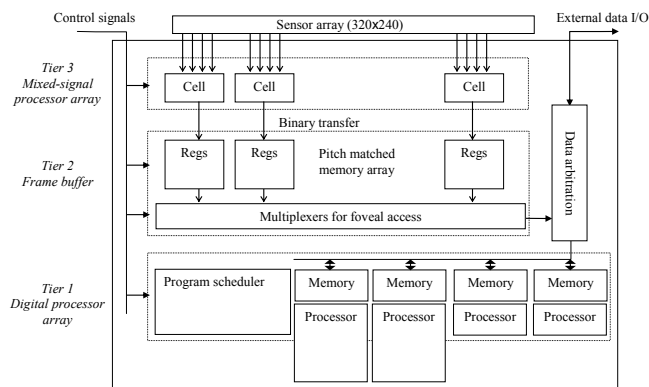


Figure 1. The architecture of the VISCUBE.

III. ALGORITHM DRIVEN PERFORMANCE

During the hierarchical feature extraction and tracking, one can find high load on the processing elements with highly regular and non-regular data transfer patterns.

The algorithm is arranged around displacement calculation of locations identified by a feature detection/selection steps. In order to calculate the required processing requirements, the amount of processors and memory needed, we have created a simulator environment with concurrent scheduling capabilities. The models cover not only functionality, but the silicon area, design complexity, and timing estimates. By executing different versions of the algorithm, we have benchmarked the options and parameters to select the final architecture. Our findings of the algorithmic driven estimates are described below.

A. Functionality

The algorithm's first step is the image capturing and storage. The second is the feature point identification. There several ways to generate candidate points, we considered a different of Gaussian (DOG) operators in combination with local extrema positions, and a Harris corner detection based solution. The former is supported by the mixed signal layer, while the later can be performed by the digital processors. The third step is the search of best match of the previously selected patterns in a new frame. This operation is a series of hierarchical, brute force, or quick (e.g. diamond) block motion estimation. And the final step is the ego-motion compensation, and further analyses. The flowchart of the algorithm and an example for the behavioral simulator's scheduling result can be seen in Fig. 2 and Fig. 3., respectively. There are intermediate steps as well, namely data transfer between the different operations. Note, that different image scales (1:2 downscaled versions in a series) are used at different steps.

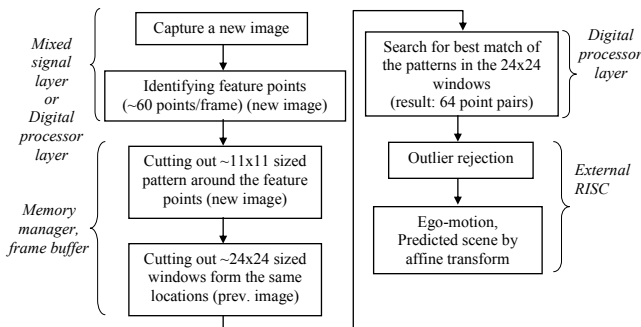


Figure 2. Flow-chart of the image registration algorithm.

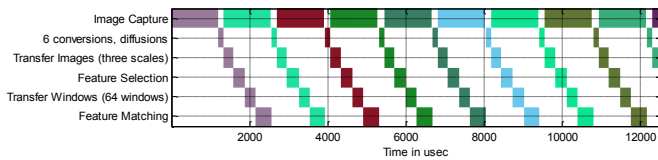


Figure 3. Scheduling example of the algorithm.

B. Hardware parameters

During the derivation of the minimally required hardware content, the calculations took into account the digital processor [3] and the mixed-signal layer capabilities. The frame, window and template sizes are used for estimating the memory requirements of the processors. Finally, the data transfer throughput was also parameterized as a function of image sizes and the selection of area or bus I/O structure. It is

worth to mention, that the design complexity choices strongly affected each other, as the given technology is an experimental one without extended industrial design kit support and off-the-shelf IP compilers and libraries (e.g. the usage of dual port memories speeds up the system by enabling parallel data transfer and processing reducing the processor number, while its nearly double size and the custom design efforts outscored their usage).

C. Selected parameters

Regarding the optimization cycles and results, we have to mention the given constraints. The most fundamental ones came from the UAV framework, namely the targeted frame rate (near 1000 fps), frames latency (1), silicon area (at most 1x1cm²), and power consumption (<1W). Counting for QVGA sized images, 24x24 feature windows and 8x8 patterns, and taking into account the transfer time, one can get an approximate processing need of 5-10 GOPS. Considering the capabilities of the used Xenon derivatives, it results in 50-100 cores (final choice is 64). The memory requirements showed two separate values. During feature point identification 0.5 Kbyte is sufficient per processor supposing the above estimated core number, while fovea processing (block matching) minimum 1.5 Kbyte per processor is required. In the final architecture 75% of the processors had got 1 Kbyte and 25% had got 2 Kbyte memory (though reduces the number of processed fovea, limited by silicon area). Table I. depicts the estimated operations counts for the given algorithm.

TABLE I. DIGITAL PROCESSOR LAYER FEATURE SEARCH AND SELECTION REQUIREMENTS

Operation	Cycles ^a $W=24, M=8, P=64,$ $L=160 \times 120,$	Time ^b
Harris corner based feature point extraction $\sim K^2 * L/P$	32k	320 μ sec
Brute force SAD: $\sim (W^2 * M^2)$	36k	450 μ sec
Diamond search: $\sim 20 * M^2$	1.2k	30 μ sec
Data collection, best selection: $\sim 20 \dots W^2$	20..0.5k	<10 μ sec
Three scale data transfer: $\sim 3 * P * W^2$	110k	110k/BW ^c

a. W = window width/height, M = mask width/height, P = number of windows/processors, L = total image size, K = gradient calculation window

b. 100 MHz core clock speed

c. BW = framebuffer – processor array I/O bandwidth (byte/sec)

Residing the feature extraction step completely into the mixed-signal layer, we found similar processing necessities. The solution at this level is the use of relatively low speed, high precision, massively parallel distributed solution (160x120 cores).

The frame buffer and the mixed-signal layer transfer rate is near 200 Gbit/seconds – due to its single slope ADC solution. This speed is easily achieved by area connection using the TSV capabilities of the technology. For the I/O need between the upper layers and the digital processor layer, considering the required data amount to be transferred is three image scales, and randomly positioned feature windows, we identified 400-500 MByte/seconds transfer rate. Surprisingly low value, and resulted in the choice of bus based interface instead of area I/O.

IV. SIGNAL, POWER, GROUND, HEAT DISTRIBUTION

In the VISCUBE implementation, there are three SOI layers with three metal layers on each, and two additional back metal layers in total.

A. Power estimation and consequences

The digital processor array consumes the most from the three layers. As front heat sink cannot be attached (the top layer is covered by optical sensors), the digital layer is placed at the bottom layer of this stack. As a result, the signal, power, ground routing required careful design, as there is no direct electrical connection to the packaging of this layer. In this section, we describe the power estimation, the inter-tier connection scheme and floorplan.

The digital processor layer has been designed by conventional 2D CAD tools. The challenge has been the long iteration cycles from modifying the RTL code to the speed, area, power, and IR drop analyses involving the iteration of the 3D connectivity floorplan.

The estimated power peak consumption of the processor array is around 450 mW@1.5V. It is important to note, that due to aggressive clock gating (>95% ratio), the switching/internal dynamic power component of peak consumption drop to 1/4th in average. Due to the significant leakage, the overall consumption drops less than this ratio. The estimated peak consumption are listed in Table II showing the effect of the clock gating.

TABLE II. POWER CONSUMPTION ESTIMATIONS

Power type	Consumption [mW@1.5V]
Peak overall	445 / 946*
Peak dynamic	163 / 662*
Leakage	282 / 284*
Memory	280

* The two values show clock gating enabled and disabled cases.

The power/ground interconnection sizes, locations, and the power routing are based on these estimates using 2D tools. At system level, the power consumption augmented with the two upper layers' operation, remains below 600 mW.

B. Floorplan

The floorplan of the digital layer is derived from the routing points connected to higher tiers. The motivation of the placement of these interconnectors was twofold: as close to the bonding pads as possible and the lack of free space within the upper layers' core area. The straightforward solution is the placement at the edge of the layers, hence enabling the most area possible for the frame buffer and processor array layer and maintain the close bonding locations. An illustration of this idea is shown in Fig. 4. A layout section of the power down via banks can be seen in Fig. 5.

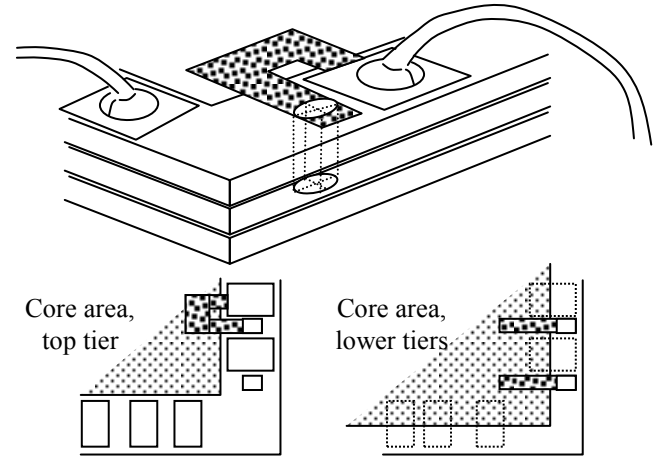


Figure 4. Illustration of the signal and power supply connection from bond wiring down to the lower tiers.

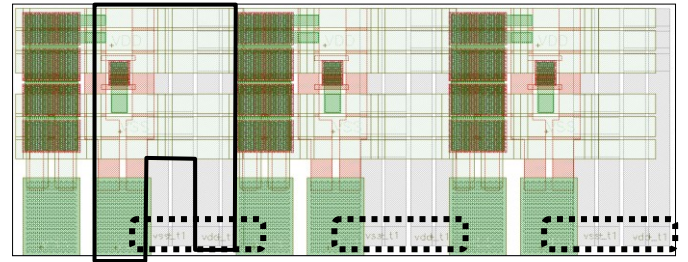


Figure 5. Layout section of the signal and power supply connection from landing pads down to the lower tiers. The dotted areas are the down via banks, while the solid U shape is the simplified pad to via bank connection.

The top level floorplan contains separated pad rings, in order to mitigate possible crosstalk between the sub-systems (analog/digital). The power and signal routing points are shown in Fig. 6. and the IR drop and heat conduction simulation driven preliminary floorplan of the digital tier can be seen in Fig. 7.

V. CONCLUSION

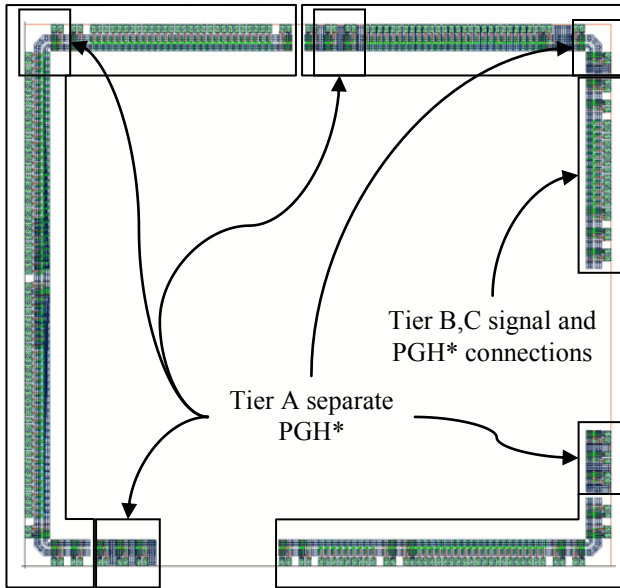
Implementation considerations are given of a 3D multi-layer complex vision system. In the final paper, in case of acceptance, we plan to give more detailed IR drop and heat distribution calculations, and the uneven memory sized array architecture as well. These details are not finalized so far, as the design is in the late design phase.

ACKNOWLEDGEMENT

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

REFERENCES

- [1] Ákos Zarándy, Dávid Fekete, Péter Földesy, Gergely Soós, Csaba Rekeczky, "Displacement calculation algorithm on a heterogenous multi-layer cellular sensor processor array", CNNA2010.
- [2] Péter Földesy, Ricardo Carmona-Galan, Ákos Zarándy, Csaba Rekeczky, Angel Rodríguez-Vázquez, Tamás Roska: „3D multi-layer vision architecture for surveillance and reconnaissance applications”, ECCTD-2009 Antalya, Turkey.
- [3] 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.



* PGH: power/ground/heat conducting inter-tier via bank

Figure 6. Signal and power supply connection points on the top tier.

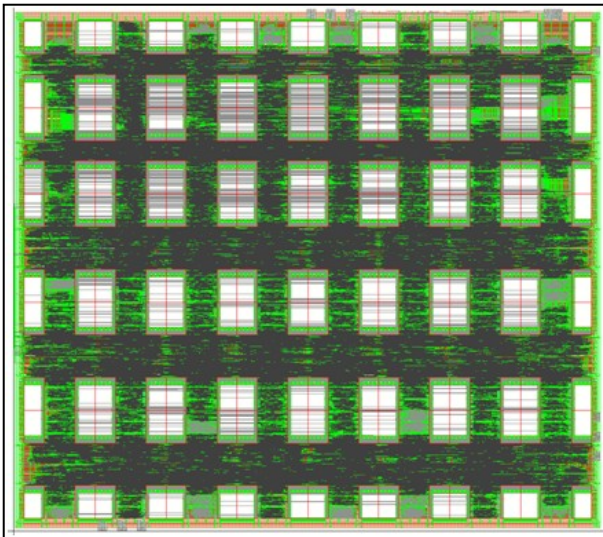


Figure 7. The floorplan of the digital processor array.

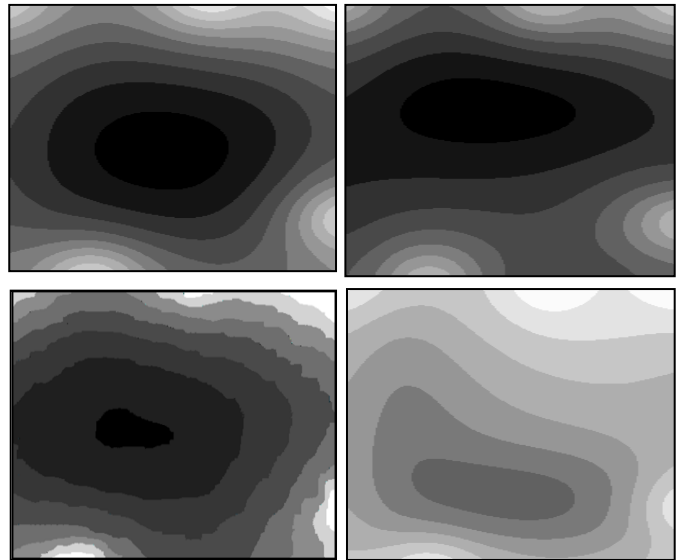


Figure 8. IR drop on core power supply – darkest region shows more than 0.2V but less than 0.25V drop at 1.5V nominal value. The lightest regions at the edges are the feed points coming from the upper layers.