16.1 iVisual: An Intelligent Visual Sensor SoC with 2790fps CMOS Image Sensor and 205GOPS/W Vision Processor

Chih-Chi Cheng[1], Chia-Hua Lin[1], Chung-Te Li[1], Samuel Chang[1], Chia-Jung Hsu[2], Liang-Gee Chen[1]

[1]National Taiwan University, Taipei, Taiwan, [2]UMC, Hsinchu, Taiwan

Visual sensors combined with video analysis algorithms can enhance applications in surveillance, healthcare, intelligent vehicle control, human-machine interfaces, etc. Hardware solutions exist for video analysis. Analog on-sensor processing solutions [1] feature image sensor integration. However, the precision loss of analog signal processing prevents those solutions from realizing complex algorithms, and they lack flexibility. Vision processors [2,3] realize high GOPS numbers by combining a processor array for parallel operations and a decision processor for other ones. Converting from parallel data in the processor array to scalar in the decision processor creates a throughput bottleneck. Parallel memory accesses also lead to high power consumption. Privacy is a critical issue in setting up visual sensors because of the danger of revealing video data from image sensors or processors. These issues exist with the above solutions because inputting or outputting video data is inevitable.

iVisual is characterized as follows: 1) Privacy is protected by integrating 2790fps CMOS Image Sensor, 76.8GOPS vision processor and 1Mb storage. It is a light-in-answer-out SoC, and no video data need to be revealed outside the chip. 2) Feature processor eliminates the throughput bottleneck and increases throughput 36%. 3) The 205GOPS/W power efficiency is 5× better than previous works [2, 3] and is achieved by introducing a feature processor, a gated-clock scheme and by reducing memory accesses.

Figure 16.1.1 shows the iVisual chip with four major parts: CMOS image sensor (CIS), global processor (GP), feature processor (FP) and decision processor (DP). GP is a parallel data in, parallel data out processor and controls the bitplane memory. FP is a parallel data-in, scalar-out processor and therefore eliminates the throughput bottleneck of data conversion. The DP processes scalar-in, scalar-out operations, that are usually decision results that further control the program execution of the GP and FP.

The CIS is frame-pipelined with GP, FP and DP to increase hardware utilization. The port of bitplane memory is shared by CIS and GP; port collision is automatically handled. The port sharing of bitplane memory reduces SRAM area 64% and die area 16% with average collision probability below 0.1%. GP, FP and DP work concurrently. For each instruction, the availability of required resources is checked, including resources in other processors. An instruction will be executed only when all required resources are available. This simple scheme ensures minimum inter-processor communication to synchronize the three processors and increases throughput 23% compared with tightly-coupled processors [2]. The clocks of unused resources are turned off to reduce power.

Figure 16.1.2 shows the CIS read-out circuits. High-gain read-out circuits have been proven to have better SNR [4]. A gain stage before the ADC with four adjustable gains is provided. For the ADC, SAR-based [5] and ramp-based architectures [6] are combined to achieve a better area-speed trade-off. Compared with the conventional SAR architecture, the required cycle count is increased from 18 cycles to 20 cycles per sample, while ADC area is reduced 48.1%. The CIS has a peak frame-rate of 2790fps because of the parallel read-out architecture.

Figure 16.1.3 shows the architecture and features of the vision processor. The GP execution unit is a SIMD processor array with 128 processing elements (PEs). The PE cache lies between the PE array and bitplane memory to reduce memory access 94%, saving 726mW of power. The PE cache itself consumes 134mW. Various bitplane memory access patterns and storage allocation schemes are provided to reduce the program size and increase storage density. To enhance flexibility, each PE is indexed and has its own conditional control. PE operations, conditional control and bitwidth control can be executed in a cycle because of the high bandwidth provided by the PE cache. Multi-resolution processing is crucial in algorithms such as face detection and object tracking. Four modes of single-cycle upsample/downsample are provided.

The FP eliminates the throughput bottleneck of data conversion from processor array to decision processor. FP is a parallel data-in, scalar-out processor that provides single-cycle feature extraction of data from the GP. The instruction set is designed from the analysis of algorithms and Intel OpenCV library. For example, the index of input data with minimum value can be extracted for calculating an object bounding box; the number of samples with value within a certain range can be extracted for color histograms. A tree-structured ALU architecture ensures a short timing path. Flexibility is increased by adding an enable signal in each input sample: calculations ignore disabled samples. The FP increases throughput 36% while occupying 4% area and consuming 5% of total power. Therefore, both area efficiency and power efficiency of iVisual are greatly increased. DP is a 32b processor with a MIPS-like instruction set and out-of-order control on parts of instructions. The DP register file is enlarged to access the data in GP. The DP can also control the program execution of GP and FP.

Figure 16.1.4 compares the throughput of iVisual and the estimated throughput of the XETAL-II architecture [2]. Two execution units exist in XETAL-II: a processor array (LPA) and a decision processor (GCP). The effectiveness of the FP is illustrated in an example of calculating the minimum value of a frame. The processor array first calculates the minimum value of each column in parallel. In XETAL-II, the GCP then has to process the data from the LPA column-by-column. This is the throughput bottleneck. With iVisual, however, FP can extract the minimum value in a cycle. The table in Fig. 16.1.4 summarizes the comparison results.

The table in Fig. 16.1.5 shows the measured throughput of commonly used operations for video analysis with 128×128 resolution. For different video resolutions, the GP can reconfigure the storage allocation to process multiple or fractional rows per cycle. High throughput is achieved by FP eliminating the throughput bottleneck and the synchronization scheme maximizing the utilization. To show the capability of iVisual when handling complex algorithms, a posture analysis algorithm is also illustrated with a flowchart shown in Fig. 16.1.5.

Figure 16.1.6 and Fig. 16.1.7 show measured chip features and the die photo, respectively. iVisual is implemented on a 7.5×9.4mm^2 die in a UMC 0.18μm 2P4M CIS process. Throughput is increased 36% with the introduction of FP and is further increased by 23% through use of the synchronization scheme. 205GOPS/W power efficiency is achieved thanks to FP, PE cache and gated-clock scheme. The comparisons of power efficiency and area efficiency are also illustrated.

Acknowledgements:
The authors thank Peter Chang, UMC University Program and ST team for the process support. This project is funded by Himax Technologies.

References:

[1] R. M. Philipp, et al., “A 128×128 33mW 30frames/s Single-Chip Stereo Imager,” *ISSCC Dig. Tech. Papers*, pp. 506-507, Feb. 2006.

[2] A. Abbo, et al., “XETAL-II: A 107 GOPS, 600mW Massively-Parallel Processor for Video Scene Analysis,” *ISSCC Dig. Tech. Papers*, pp. 270-271, Feb. 2007.

[3] S. Kyo, et al., “A 51.2 GOPS Scalable Video Recognition Processor for Intelligent Cruise Control Based on a Linear Array of 128 4-way VLIW Processing Elements,” *ISSCC Dig. Tech. Papers*, pp. 48-49, Feb. 2003.

[4] N. Kawai and S. Kawahito, “Noise Analysis of High-Gain, Low-Noise Column Readout Circuits for CMOS Image Sensors,” *IEEE T. Electron Devices*, vol. 51, no. 4, pp. 185-194, Feb., 2004.

[5] I. Takayanagi, et al., “A 1.25-inch 60-frames/s 8.3-M-Pixel Digital-Output CMOS Image Sensor,” *IEEE J. Solid-State Circuits*, vol. 40, pp. 2305-2314, Nov., 2005.

[6] S. Yoshihara, et al., “A 1/1.8-inch 6.4MPixel 60 frames/s CMOS Image Sensor with Seamless Mode Change,” *ISSCC Dig. Tech. Papers*, pp. 492-493, Feb. 2006.

978-1-4244-2010-0/08/$25.00 ©2008 IEEE

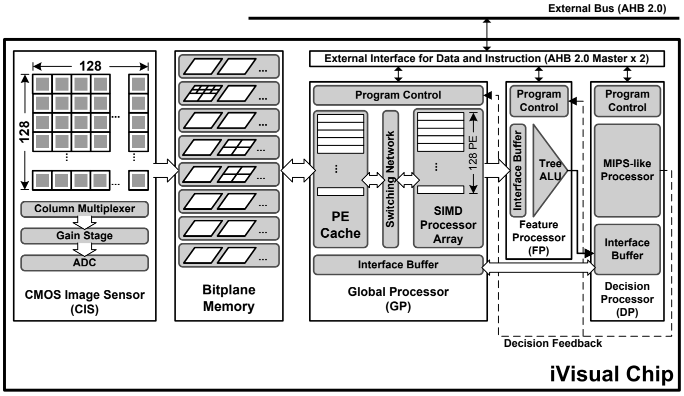


Figure 16.1.1: The iVisual hardware architecture.

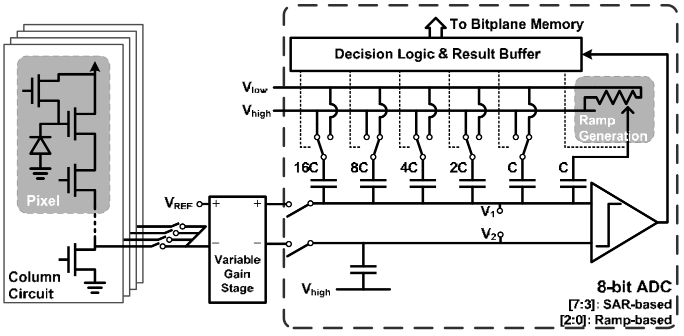


Figure 16.1.2: The CIS read-out circuits.

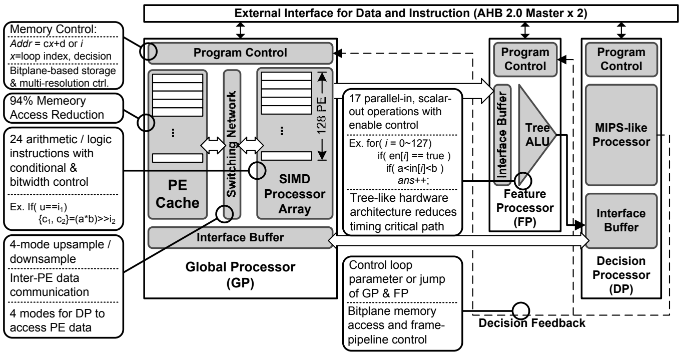
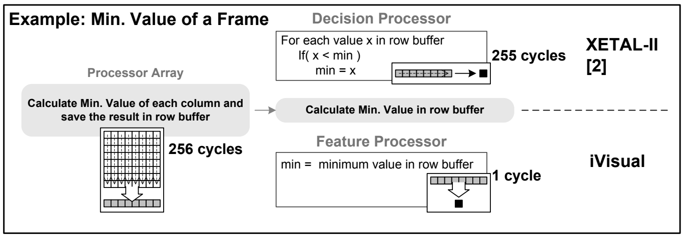

External Interface for Data and Instruction (AHB 2.0 Master x 2)
Memory Control:
Addr = cx+d or i
x=loop index, decision
Bitplane-based storage & multi-resolution ctrl.
94% Memeory Access Reduction
24 arithmetic / logic instructions with conditional & bitwidth control
Ex. If(u==i1) {c1, c2}=(a*b)>>i2
4-mode upsample / downsample
Inter-PE data communication
4 modes for DP to access PE data
Program Control
PE Cache
Switching Network
SIMD Processor Array
128 PE
Interface Buffer
Global Processor (GP)
17 parallel-in, scalar-out operations with enable control
Ex. for(i = 0~127) if(en[i] == true) if(a<in[i]<b) ans++;
Tree-like hardware architecture reduces timing critical path
Control loop parameter or jump of GP & FP
Bitplane memory access and frame-pipeline control
Program Control
Interface Buffer
Tree ALU
Feature Processor (FP)
Program Control
MIPS-like Processor
Interface Buffer
Decision Processor (DP)
Decision Feedback

Figure 16.1.3: Architecture and features of vision processor.


Example: Min. Value of a Frame
Decision Processor
For each value x in row buffer
If(x < min)
min = x
255 cycles
XETAL-II [2]
Processor Array
Calculate Min. Value of each column and save the result in row buffer
Calculate Min. Value in row buffer
256 cycles
Feature Processor
min = minimum value in row buffer
1 cycle
iVisual

Operation Description	Throughput (Cycles/frame)		Cycle Reduction Ratio
	Xetal-II (estimated)	iVisual	
Connect Component Extraction	131072	55370	57%
Image Mean	257	130	50%
16-bin Intensity Histogram	8193	2050	75%
Min/max Value of Frame	511	257	50%
Subsample of Image	6240	256	96%
Object Bounding Box Calculation	1924	903	53%
Elliptical Matching	17759	12127	32%
7-Object Segmentation + Tracking	203464	123774	39%

Note: **1. XETAL-II architecture model: PE array, 32-b processor, flag-based addition & selection**

2. The number of PE in processor array in XETAL-II is scaled to be the same as iVisual

3. Working frequency: 84MHz in XETAL-II (90nm process), 50MHz in iVisual (.18um process)

Figure 16.1.4: Architectural comparison with XETAL-II assuming 128×128 frame size.

Performance for Single Operations

Operation Description	Throughput
Frame-Level Processing	
3x3 2-D FIR Filter	12204fps (0.25cycle/pixel)
Sobel Image Gradient	55741fps (0.05cycle/pixel)
Harris Corner Detector	1367fps (2.2cycle/pixel)
Erosion (Dilation)	55741fps (0.05cycle/pixel)
3x3 Median Filtering	6300fps (0.48cycle/pixel)
Row-Based Pipeline among Three Processors	
6-Dimensional (Affine Model) Motion Estimation	384fps (7.95cycle/pixel)
Object Bounding Box Extraction	48449fps (0.06cycle/pixel)
2-D Projective Histogram	97465fps (0.03cycle/pixel)
Histogram Equalization	43327fps (0.07cycle/pixel)
16-Bin Intensity Histogram	22946fps (0.13cycle/pixel)
Frame-Based Pipeline among Three Processors	
Elliptical Matching of Object	4123fps (0.74cycle/pixel)
Connected Component Calculation	903fps (3.38cycle/pixel)
Horn Optical Flow Calculation	583fps (5.23cycle/pixel)
Integral Image on 3 Resolutions	3460fps (0.88cycle/pixel)
Object Area Extraction	128865fps (0.02cycle/pixel)

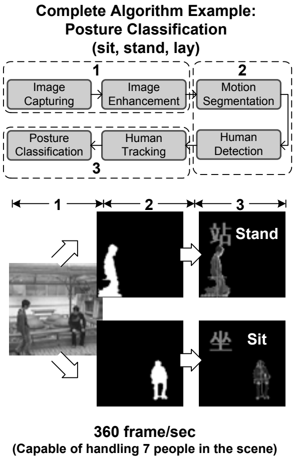
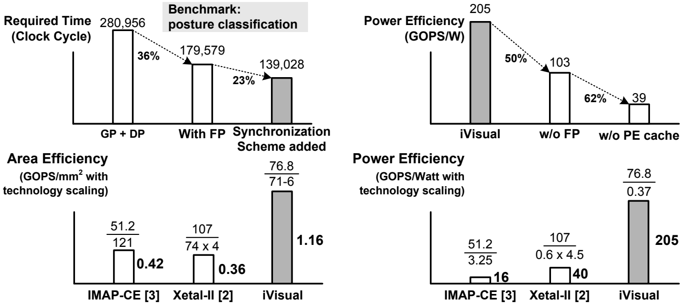


Figure 16.1.5: Measured throughput of iVisual.

Technology	UMC 0.18µm 2P4M CMOS Image Sensor Process	Internal Buffer	1Mb
Core Area	7.5mm x 9.4mm	Power Consumption	CIS: 81mW Vision Processor: 374mW
Package	BGA 256 pin	Peak Performance	76.8GOPS
Operating Frequency	50MHz	External Interface	Two AHB 2.0 Masters
Operating Voltage	2.1V	Pixel Type	3T Pixel Cell
Temperature	25ºC	CIS Resolution	128x128



technology scaling of power (90nm to 180nm): $P_{180}=P_{90}\times(C_{180}/C_{90})\times(V_{180}/V_{90})^2=P_{90}\times2\times(1.8/1.2)^2=P_{90}\times4.5$

Figure 16.1.6: Measured chip features.

Continued on Page 615

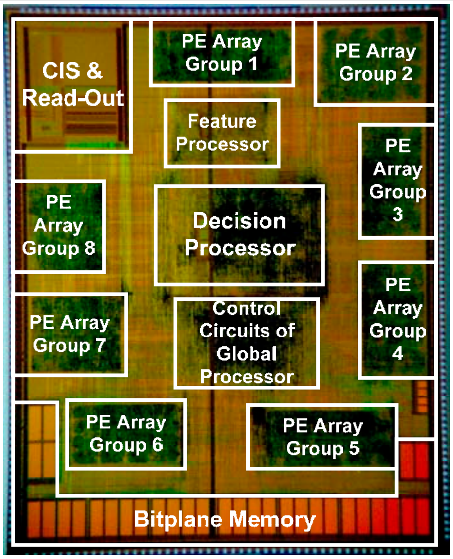


Figure 16.1.7: Chip micrograph.