FaceDET—a 1.4 GOPS Face Detection System-on-chip based on Binary Decision Trees

E. Azarkhish, E. Türetken, P. Nussbaum, S. Emery

FaceDET is a high performance low-power system-on-chip (SoC) designed to perform face detection in edge-applications with power requirements of a few milliwatts. The design follows a modular approach with a general-purpose RISC-V based processor (main controller), a memory-mapped Binary Decision Tree (BDT) accelerator (BDTAcc), and the associated memory subsystem and peripherals. Synthesis results prove feasibility with system frequency reaching around 180 MHz, providing nominal system performance up to 1.4 GOPS [1] and a memory bandwidth above 20 GB/s.The place-and-route of the complete system-on-chip is being done in the Globalfoundries 22 nm FDX technology.

Figure 1 illustrates an overview of the FaceDET SoC. It reuses icyflex-V, a 32-b RISC-V based micro-processor designed and maintained by CSEM, as its main controller. This is to enhance programmability and ensure compatibility with the software stacks developed inside CSEM and by the rest of the RISC-V community. 16 KB of ROM and 64 KB of private memory (L0) are dedicated to the main-controller, accessible by icyflex-V and its peripherals.

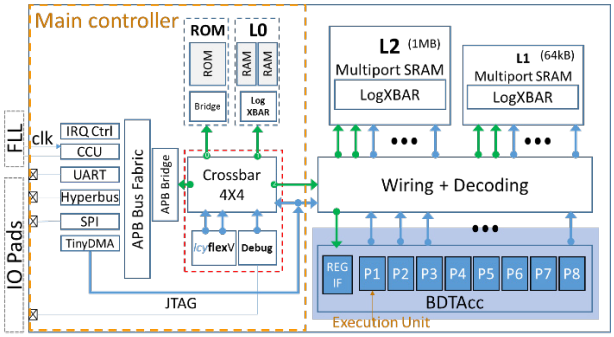

Figure 1: An overview of the FaceDET SoC.

The icyflex-V core handles communication with the peripherals, including CSEM's low-power VGA imager [2] attached to the Hyperbus, an external flash attached to standard SPI, an FLL for clock generation, and a UART for user interaction. The main controller has a visibility of the whole memory space including all memory levels, peripherals, and the accelerator's registers. The Central Configuration Unit (CCU) is responsible for configuring pads, peripherals and memories, as well as, power management.

An optimized fetch-and-deposit DMA (TinyDMA), with a very small footprint and power consumption, is responsible for transfers between memory and the peripherals. The Hyperbus IP enables high-speed communication (>100 MBps) with the imager.

The key differentiating point about the architecture of FaceDET is a highly parallel memory organization with a unified address space thanks to an especially designed logarithmic crossbar interconnect (LogXBAR), providing all-to-all connectivity with 2-by-2 arbitration trees. Word-level interleaving is adopted to remove the burden of bank-conflicts from the programmers, and banking factor has been increased to 4, to naturally reduce bank conflicts to below 5%. This is in opposition to standard bank-interleaving schemes, in which the programmer must explicitly allocate memory banks to accelerator units and mind the bank conflicts. The L1 (64 KB) and L2 (1 MB) memories, both organized in 16 banks, operate at the same system frequency, except that access to L1 is less costly in terms of energy compared to L2. Efficient use of the L1 memory by the software helps bring down the overall energy consumption.

BDTAcc implements an optimized version of the AdaBoost algorithm using 8 parallel execution units (ExU). Each ExU contains a hardware state-machine, capable of traversing binary decision tree nodes in the memory and making decisions independently. AdaBoost is an easily parallelizable algorithm breaking the image into several overlapping bounding-boxes of different sizes and orientations and calculating the likelihood of detecting faces in them with two levels of tree based classifiers, as illustrated in Figure 2. A large batch of bounding boxes can be dispatched to each ExU to parallelize and accelerate execution.

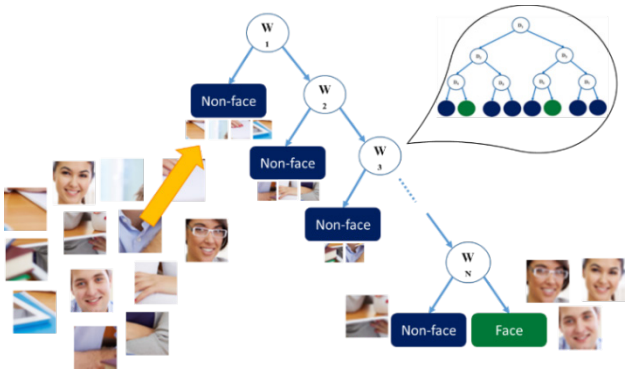

Figure 2: An illustration of the AdaBoost face-detection algorithm.

Each ExU implements two variations of AdaBoost: FAST with a larger memory footprint but fewer memory accesses per operations, and COMPACT with smaller footprint and more accesses resulting in higher consumption and execution time. The user program can define a threshold to switch from the FAST to the COMPACT method to dynamically benefit from both of them.

The 22 nm FDX technology from GlobalFoundries has been selected as one of the most advanced nodes for the IoT and low-power applications. We use body-biasing to dynamically adjust face-detection performance in a trade-off with power consumption.

Synthesis results show that more than 90% of the area is devoted to the L1 and L2 memories and their interconnections. BDTAcc uses less than 2%, and the rest of the area is used by the main controller. A new revision is under preparation with a convolutional neural network (CNN) accelerator beside the BDTAcc, allowing for hierarchical detection/classification with the two approaches to find the right balance between accuracy and energy consumption.

[1] Comparison of two integers is considered an operation.

[2] P.-F. Rüedi, *et al.*, "An Ultra-low-power High Dynamic Range Image Sensor", CSEM Scientific and Technical Report (2018) 101.