The visual attention mechanism, which is the way humans perform object recognition [1], was applied to the implementation of a high performance object recognition chip [2]. Even though the previous chip achieved 50% gain of computational cost [2], it could recognize only one object in a frame so that it is not suitable for advanced multi-object recognition applications such as video surveillance, intelligent robots, and autonomous vehicle navigation [3].

A real-time multi-object recognition processor is presented based on the bio-inspired visual perception algorithm. The proposed recognition processor has 4 features: 1) 3-stage pipelining with grid-based region-of-interest (ROI) processing for high recognition rate, 2) Neural perception engine (NPE) with three bio-inspired neural and fuzzy processing units for multi-object perception and segmentation, 3) Low latency multi-castable Network-on-Chip (NoC) for high bandwidth integration platform, and 4) Workload-aware power management for low power consumption.

Figure 8.3.1 shows the overall block diagram of the proposed processor. It is composed of 21 IPs on a chip: a NPE, a SPU task/power manager (STM), 16 SIMD processor units (SPUs), a decision processor (DP), and 2 external memory interfaces. The bio-inspired NPE is composed of motion estimator (ME), visual attention engine (VAE2), and object segmentation engine (OSE). It performs global feature extraction and object segmentation using the neural and fuzzy processing to extract ROIs. The 16 SPUs perform complex and data-intensive image processing for the selected ROIs. The detailed block diagram of the SPU is shown in Fig. 8.3.2. Each SPU consists of eight 16b SIMD processing elements (PEs), 1 scalar datapath, 12KB 128b wide data memory with 2 aligners, and 2 DMA. Dual-issue VLIW enables parallel execution of data processing and data transfer operations. A register file with 5-read and 3-write ports enables PE to execute 2-way 8b multiply-and-accumulate, 3-operand 16b min/max compare, and 32b accumulation in a single cycle. For low power consumption, the 16 SPUs are divided into 4 voltage/frequency domains called SPU cluster (SPC). The STM dynamically assigns ROI tasks to 16 SPUs, and controls 4 SPC power domains. The DP recognizes each object using the database search for the generated descriptor vectors by the SPUs.

Figure 8.3.3 shows the 3-stage multi-object recognition with grid-based ROI processing. It is composed of: 1) visual perception, 2) descriptor vector generation, and 3) object decision. The visual perception stage classifies the boundaries of the multiple objects based on the extracted static and dynamic features from the input images. It extracts the ROIs for each object in a 40x40 pixel tile. Extraction of the ROIs in the visual perception stage reduces the workload of the following stages by focusing their operations on only the extracted ROIs. The descriptor vector generation stage calculates descriptor vectors for the selected ROIs. Then, in the object decision stage, the descriptor vectors of the objects are recognized through iterative matching with the database. In the proposed processor, the 3 stages of the object recognition pipeline are directly mapped to the NPE, 16 SPUs, and DP, respectively. The STM controls the processing speed of 16 SPUs according to the workloads coming from the NPE to match the processing time of each stage of the pipeline. As a result, task pipelining with the grid-based ROI processing reduces computation area by 41%, and achieves a 3.8x performance improvement compared to the previous serial object recognition based on column-wise processing [2].

The NPE consists of a 32b RISC processor, cellular neural network based VAE2 [4], ME, OSE, and 24KB shared memory. After the VAE2 and ME generate a saliency map from the 1/8 down-sized 80x60 input images, the OSE finally determines the ROIs for each object by selecting the 10 most salient points and by growing the regions starting from the selected 10 seeds.

The STM performs workload-aware power domain management and IP-level clock gating as shown in Fig. 8.3.6 to reduce the power consumption of the 16 SPUs. The STM determines the number of activated SPC power domains by measuring the per-frame workload from the NPE, and dynamically assigns ROI tasks to the activated SPCs. In the activated power domains, the clock of each SPU is gated by a software request for a further reduction of SPU dynamic power. Through the domain management and the clock gating, the power dissipation of 16 SPUs is reduced by 38% at a 60fps sustained frame rate.

Figure 8.3.7 shows the chip micrograph and summarizes its features. It is fabricated in 0.13μm CMOS technology and occupies 56mm² containing 36.4M transistors with 3.73m gates and 396KB on-chip SRAM. The 1.2V processor achieves 60fps object recognition for a maximum of 10 objects with 496mW power consumption at 200MHz IP clock and 400MHz NoC clock frequency. Its 296GOPS/W power efficiency is the highest among the previously reported parallel processors [2, 7, 8] as shown in Fig. 8.3.7.

References:
Please click on paper title to view Visual Supplement.
Figure 8.3.7: Chip micrograph and summary.