A 1.2mW On-Line Learning Mixed Mode Intelligent Inference Engine for Robust Object Recognition

Jinwook Oh, Seungjin Lee, Minsu Kim, Joonsoo Kwon, Junyoung Park, Joo-Young Kim and Hoi-Jun Yoo
Department of EE, KAIST, Yuseong, Daejeon, Republic of Korea
E-mail: jinwook@eeinfo.kaist.ac.kr

Abstract

An intelligent inference engine (IIE) is proposed as a controller for low power high speed robust object recognition processor. It contains analog digital mixed mode neuro-fuzzy circuits for the on-line learning to increase attention efficiency. It is implemented in 0.13um CMOS process and achieves 1.2mW power consumption with 94% average classification accuracy within 1us operation. The 0.765mm^2 IIE achieves 76% attention efficiency, and reduces power and processing delay of the 50mm^2 recognition processor by up to 37% and 28%, respectively, with 96 % recognition accuracy.

Introduction

The visual attention that determines regions-of-interest (ROIs) with which human can focus his view on a specific region is a very efficient pre-processing procedure for computer vision, especially real-time object recognition [1]. But high recognition rate cannot be obtained with bottom-up attention alone. The state-of-the-art object recognition processor [2] adds a human-like top-down feedback loop in addition to bottom-up attention as shown in Fig.1. By obtaining more precise ROIs, the enhanced attention processing provides not only robust recognition but also processing speed and power consumption gain.

In this paper, an intelligent inference engine (IIE) is proposed to realize the top-down attention path. The versatile adaptive neuro-fuzzy inference system (VANFIS) and its on-line learning turn out high classification accuracy. In addition, it is implemented by combining analog neuro-fuzzy circuits and a digital controller to reduce the area and power consumption with high speed operation, and its detail architecture will be explained in the next section.

On-line Learning and Mixed Mode Architecture

The parameter perturbation algorithm [4] for NFL is known for fast and simple hardware implementation with its low computation complexity. Among perturbation algorithms, according to simulation results of Fig. 2, sequential perturbation scheme is good for learning accuracy, and simultaneous perturbation scheme is advantageous in high speed. In this design, we adopt the adaptive perturbation algorithm (APA) that dynamically chooses only the advantageous one by tracing the error saturation point. Thanks to APA, the IIE can achieve 94% classification accuracy for 25 database objects.

Fig. 3 shows the architecture of the proposed IIE. It is composed of an analog VANFIS core which is hardware efficient due to parallel and collective analog NFL [5], and a digital controller consisting of a 16-bit RISC processor, a learning accelerator, and a 4KB parameter cache performs overall data control and on-line learning for VANFIS core. The two domains are interfaced by DACs/ADC. The learning accelerator accelerates the proposed APA by 21% with a true random bit generation (TRBG) and a data aligner converting 16-bit digital learning results to 5-bit analog parameters.

The current mode analog circuits implement the VANFIS core composed of 5-layer Takagi-Sugeno NFL [3]: fuzzy membership function (MF), fuzzy rule base (FR), normalization, weight multiplication, and integration layer. It receives scale, orientation, and motion as inputs and generates the object confidence through the neuro-fuzzy inference. The 9 MFs characterize the 3 input features to membership grades and the FR layer generates 27 fuzzy rule compositions using min-max functions. After normalization and weight multiplication, the final output confidence is generated by simple integration. In addition, versatile granular architecture (VGA) is adopted in the MF layer and the FR layer as depicted in Fig. 3. VGA configures the number of input features and their connecting MFs for various types of object classifications. As a result, the 4 possible VGA configurations of the IIE can sustain classification accuracy more than 90% for different object classification problems.

Detailed Circuit Implementations

The two significant circuits in analog VANFIS core are various shaping MF circuits and perturbation-aided current steering DAC circuits. Fig. 4 shows the proposed controllable MF circuit that can adjust its slope, width and location. Two symmetric MOS differential pairs of M_1~M_4 and the current complementary extractor M_5~M_8 are implemented with the MOS switched arrays and two reference voltages to modify the g_m and high/low boundaries of MF. Including basic Gaussian function, this circuit can provides 5 distinctive shapes of function such as trapezoidal, triangular, s-shape and z-shape by combining three parameters. Using this extensive controllability of the MF, the VANFIS can support several types of input features with high classification performance. The parameters of MF are adjusted by the proposed DAC whose output voltages are digitally perturbed in conversion process. As shown in Fig. 5, the perturbation-aided current steering DAC has extra transistors which generate the fixed perturbation *c* on parameters by utilizing the TRBG circuits. The TRBG circuits consist of 4 meta-stable latches with control switches. The sign signal of perturbation Pconf is generated from 4 latches to compensate the mismatch of the 2 inverters with the help of TRBG controller in digital domain. Because the perturbation is smaller than the least significant bit (LSB) of DAC, the on-line learning algorithm can be conducted more sophisticatedly and rapidly than the system with digital or analog stand-alone perturbation circuits that require extra perturbation generation procedure. For weight multiplication layer the DAC also can be employed with I_{unit} replaced as the input operand current, then negative and positive lines are allocated for the sign bit b_{SIG} with a current subtractor generating the integrated current I_{DIF} and a voltage sign bit V_{SGN} as shown in Fig. 5. Compared to the equivalent digital implementation [6], the analog VANFIS core only consumes 54% area and 15% power within same processing time.

Implementation Results and Conclusions

Fig. 6 shows the chip photograph and implementation results of the proposed IIE. Implemented in 0.13um CMOS technology, the IIE achieve 1.2mW power at 1.2V and 0.765mm^2 area. As inference accelerator of a 50mm^2 robust-object recognition processor running at 200MHz, the analog VANFIS core is measured to have ~1us operation speed, which achieves 1M fuzzy logic inference per second (FLIPS) and maximum 54M connection updates per second (CUPS) as neuro-fuzzy system. Regarding the overall object recognition SoC, the attention efficiency as a performance metric is defined by n(RCT)/n(ROI), where n(RCT) is the number of ROIs containing target objects, and n(ROI) is the number of ROIs to be processed. It indicates how accurately the attention is focused on to the meaningful target area on the processed image. The fabricated IIE achieves 76% attention efficiency resulting in less than 30% of ROIs, with reduction of power consumption and processing time by 37% and 28%, respectively.

Fig. 7 shows the measurement waveforms and experimental results of IIE for on-line machine learning. For simple object features, the previous 60% confidence level is increased to about 95% within only 70us of repetitive learning sequence. In the pattern recognition processor, it achieves real-time adaptation for new object feature patterns such as occluded objects and novel viewed objects. When applying off-line and on-line learnings with APA to object tracking, the initial objects are recognized by both schemes. However, when severe occlusion and novel view change in appearance occur, the off-line stand-alone learning fails to recognize the target objects with ROIs covering whole image frame. On the other hand, the on-line learning with APA robustly recognizes each ~70% occluded and ~50° novel viewed object in 640x480 video images. Consequently, the on-line learning adaptation enables 96% recognition accuracy of the processor, which is 11% improvement compared to the off-line stand-alone learning. By adopting 0.765mm^2 IIE, the 50mm^2 object recognition processor consumes 345mW saving 37% of its power with 96% accuracy increasing it by more than 11% compared with the previous processor. As a result, it achieves 8.5mJ/frame energy efficiency for 30fps VGA video streaming in real-time operation.

References

[1] J.Y. Kim, et al., "A 201.4GOPS 496mW Real-Time Multi-Object Recognition Processor with Bio-Inspired Neural Perception Engine," in IEEE ISSCC, pp.150-151, Feb., 2009.

[2] S. Lee, et al., "A 345mW Heterogeneous Many-core Processor with an Intelligent Inference Engine for Robust Object Recognition," in IEEE ISSCC, Feb., 2010. Session 18.4

[3] J. –S.R. Jang, "ANFIS: Adaptive-network-based fuzzy inference system," in IEEE Transactions on Systems, Man, and Cybernetics, vol. 23, no.3, pp.665-685, May 1993.

[4] M. Jabri, "Weight Perturbation: An Optimal Architecture and Learning Technique for Analog VLSI Feedforward and Recurrent Multilayer Network," in Neural Computation, vol. 3, no. 4, pp. 546 - 564, May, 1991.

[5] R. Sarpeshkar, "Analog Versus Digital: Extrapolating from Electronics to Neurobiology," in Neural Computation, vol. 10, pp. 1601-1038, Mar. 1998.

[6] M. Kim, et al., "A 22.8 GOPS 2.83mW Neuro-fuzzy Object Detection Engine for Fast Multi-object Recognition," in IEEE Symposium on VLSI. Jun., 2009.

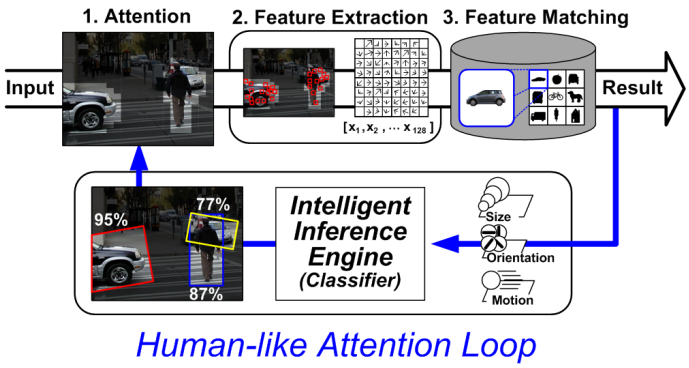


Fig.1. Human-like attention with feedback loop

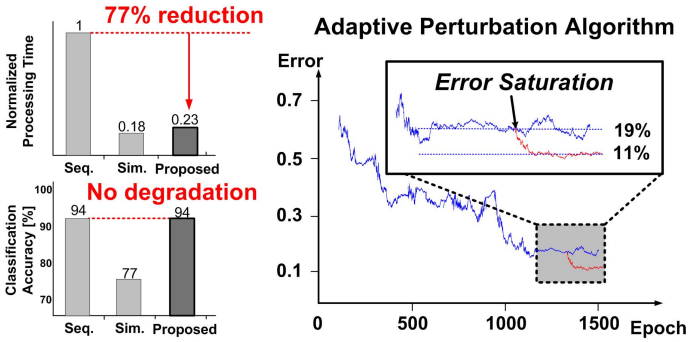


Fig.2. Proposed adaptive parameter perturbation

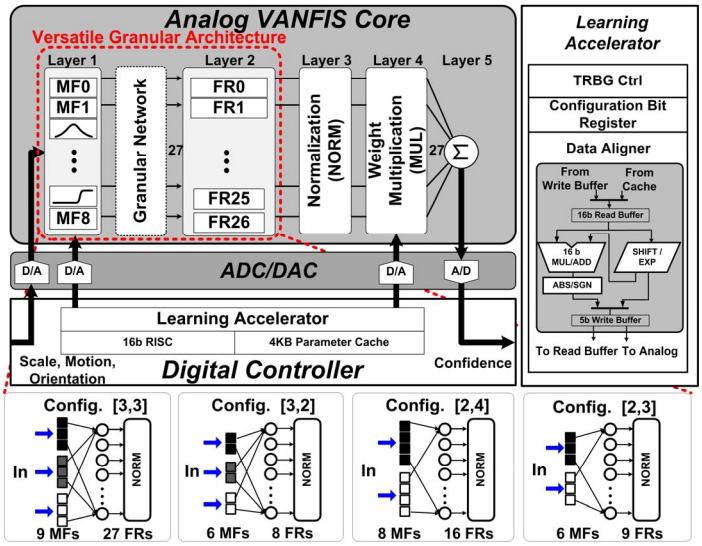


Fig.3. Intelligent inference engine architecture

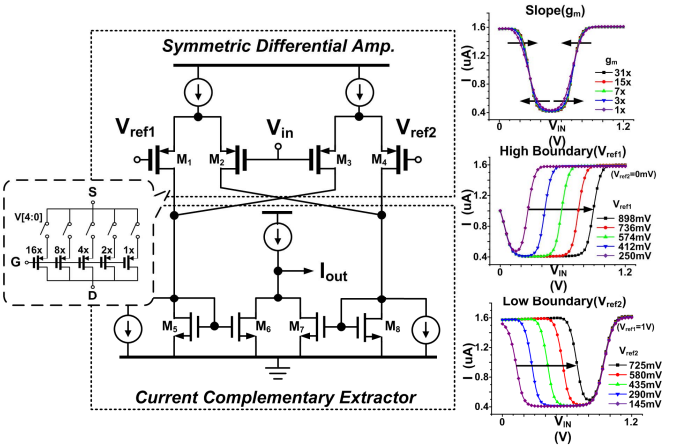


Fig. 4. Controllable membership function circuit

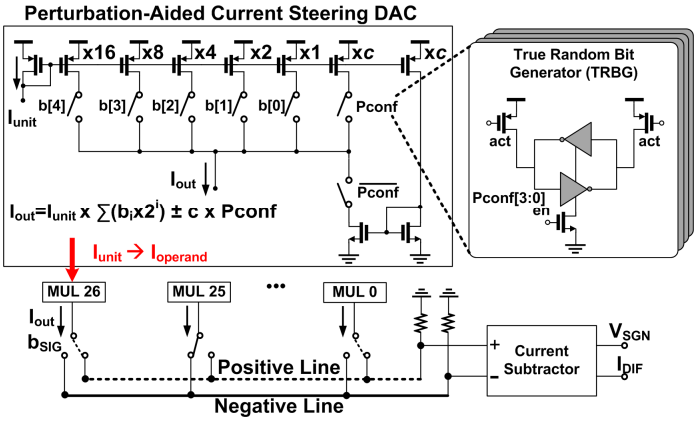
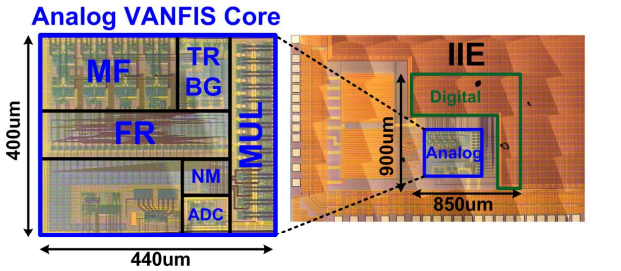


Fig. 5. Perturbation-aided current steering DAC circuits



Item	Specification	
Process Technology	0.13um 1P 8M CMOS Technology	
Area	440umx400um(Analog) / 850umx900um	
Power Supply	1.2 V for core	
Operating Frequency	200MHz for Digital Domain	
Peak Performance	Learning	MAX. 54MCUPS
	Inference	1MFLIPS(27 Rules)
Power Consumption	1.2mW	

Fig.6. Chip photograph and implementation results

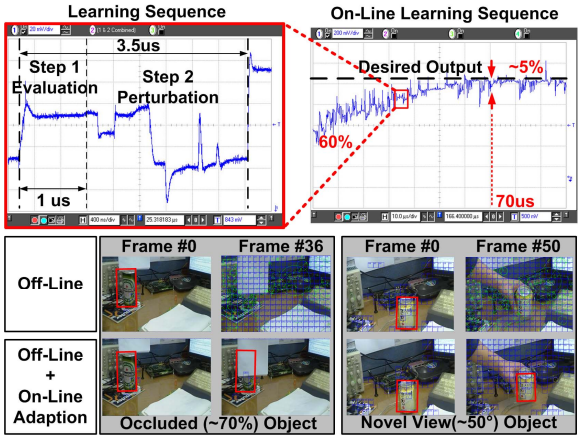


Fig.7. Experiment results of on-line learning IIE