Fast and efficient multi-layer CNN-UM emulator using FPGA

Zoltán Nagy

A Cellular Neural Network is a non-linear dynamic processor array. Its extended version the CNN Universal Machine (CNN-UM) was invented in 1993. [1] The main application area of this architecture is 2D signal or image processing. The most effective implementation of the CNN-UM architecture seems to be analog VLSI. The latest analogue CNN chip has a 128 × 128 pixel resolution and its equivalent computing power is 4 tera operation/second but its computational precision is about 7 or 8 bits. [2] In many applications these parameters are not high enough. If the resolution is higher we don't need to slice the images. If the precision is higher, less robust or more sophisticated analogical algorithms can be used.

A multi-layer CNN array can be used to solve the state equation of complex dynamical system. Currently the only method to solve the state equation of multi-layer CNN array is software simulation. If every layer has different time constant very small simulation stepsize must be chosen, thus software simulation is very slow. To achieve affordable runtimes the simulations have to be accelerated. This motivation came from the analysis of a retina model, where the retina is modelled with 3-layer CNN array and every layer has different time constant. [3]

The Falcon emulated digital CNN-UM chip was designed to reach these goals. [4] Especially flexible emulated digital CNN-UM was developed where the accuracy, template size, cell array size and the number of layers can be configured. Simulation runtimes can be hundred times shorter using the Falcon processor array compared to the software simulation. This paper describes the synthesis, implementation and optimization methods used to implement the Falcon processor array on FPGA. The Distributed Arithmetic technique is used to optimize the architecture on FPGAs. [5,6] Using this technique, smaller and faster arithmetic units can be designed than the conventional approach, where multiplier cores and adder trees are used to compute the state equation of the CNN array.

The Falcon architecture was implemented on our prototyping board, using a Virtex-300 FPGA from Xilinx Inc. The performance of the architecture was encouraging, even in a single processor configuration 20 fold speedup can be achieved compared to 1GHz Pentium III Xeon processor. The processor runs only on 80MHz clock frequency because of the limitations of the prototyping board. Using faster memories, higher speed grade FPGA or using the more advanced Virtex-E and Virtex-II FPGAs twenty times higher performance can be easily achieved. The easy scalability of the Falcon architecture makes possible to connect the processor cores in a square grid and achieve even more performance. Using re-configurable devices to implement the Falcon architecture provide us more flexibility compared to the conventional emulated digital architectures e.g. different configurations can be used on the same hardware and extra design effort is not required to implement it.

References

[1] T. Roska and L. O. Chua, "The CNN Universal Machine. An analogic array computer" IEEE Trans. On Circuits and Systems-II, vol. 40, pp. 163-173, 1993.

[2] G. Lińán, R. Domínguez-Castro, S. Espejo, A. Rodríguez-Vázquez "ACE16k: A Programmable Focal Plane Vision Processor with 128x128 Resolution" in Proc. of the 15th European Conference on Circuit Theory and Design, vol. 1, pp. 345- 348, 2001.

[3] D. Bálya, B. Roska, T. Roska and F.S. Werblin, "A CNN Framework for Modeling Parallel Processing in a Mammalian Retina", Int. J. on Circuit Theory and Applications, vol. 29, No. 3, 2002.

[4] Z. Nagy, P. Szolgay, " Configurable multi-layer CNN-UM emulator on FPGA", to be published on CNNA 2002, Frankfurt, 2002.

[5] Peled and B. Liu, “A New Hardware Realization of Digital Filters”, IEEE Trans. on Acoust., Speech, Signal Processing, vol. ASSP-22, pp. 456-462, Dec. 1974.

[6] Mintzer, L. “FIR filters with the Xilinx FPGA” in Proc. of FPGA ’92 ACM/SIGDA Workshop on FPGAs, pp. 129-134, 1992