

High Performance Image Processing by Brain Emulation

P. Luciano, *Student Member IEEE*, C.-L. Sotiropoulou, *Member IEEE*, S. Gkaitatzis, M. Viti, S. Citraro, A. Retico, P. Giannetti and M. Dell' Orso

Abstract— We present an innovative and high performance embedded system for real-time pattern matching. The design uses Field Programmable Gate Arrays (FPGAs) and the powerful Associative Memory chip (an ASIC) to achieve real-time performance. The system works as a contour identifier able to extract the salient features of an image. It is based on the principles of cognitive image processing, which means that it executes fast pattern matching and data reduction mimicking the operation of the human brain.

P. Luciano is with the University of Cassino and Southern Lazio (email: pierluigiluciano@pi.infn.it).

C.-L. Sotiropoulou, S. Citraro and M. Dell' Orso are with the University of Pisa and INFN Pisa Section (email: c.sotiropoulou@cern.ch, saverio.citraro@pi.infn.it, mauro.dellorso@pi.infn.it)

S. Gkaitatzis is with the Department of Physics of the Aristotle University of Thessaloniki (email: stamatis.gkaitatzis@cern.ch)

M. Viti is with the Department of Informatics of the University of Pisa, (email: vitimario2@gmail.com)

A. Retico and P. Giannetti are with INFN Pisa Section (email: alessandra.retico@pi.infn.it, paola.giannetti@pi.infn.it).

I. INTRODUCTION

We have built an Associative Memory (AM) system for the Fast Tracker (FTK) processor [1], a recently approved upgrade for the ATLAS trigger [2]. FTK is a high-performance embedded system based on the combination of two innovative technologies: powerful and flexible FPGAs working with standard-cell ASICs, the Associative Memory (AM) chips [3], for utmost gate integration density and maximum performance to execute the pattern matching algorithm. The most interesting processes generated at LHC are very rare and hidden in an extremely high level of background. Implementing the most powerful selections in real-time (trigger) is therefore essential to fully exploit the physics potential of experiments where only a very limited fraction of the produced data can be recorded. This is a specific case of “Big Data” problem whose solution is based on the organization of the trigger in different levels of selections [4]. At low level we exploit parallelized, dedicated hardware for an extremely efficient preprocessing step.

This trigger organization is similar to models of the vision processing task performed by the brain. Our embedded system can accelerate neurophysiologic studies of the brain. The most convincing models about brain functioning hypotheses are extremely similar to the real time architectures developed for high energy physics. A multilevel model seems appropriate to describe the brain image processing [5]: “the brain works by dramatically reducing input information by selecting for higher-level processing and long-term storage only those input data that match a particular set of memorized patterns. The double constraint of finite computing power and finite output bandwidth determines to a large extent what type of information is found to be *meaningful* or *relevant* and becomes

part of higher level processing and longer-term memory”. The AM pattern matching process has demonstrated to be able to play a key role in high rate filtering/data-reduction tasks. Simulations [5] have shown the potential of the pattern matching algorithm on static 2-D images. Since the needed computational time causes serious limits to the capability to extend these studies to 3-D images and movies, we are developing an implementation that will use the AM system [6] based on the AM chip [7] for a real-time pattern selection/filtering of the same type studied in these models of human vision. These studies could have an impact in the area of medical imaging for real-time diagnosis or any area where pattern matching is relevant and computing is a limiting factor.

II. THE FILTERING ALGORITHM

Fig. 1 shows the results of the simulations of the model described in [5] where pattern matching with relevant patterns is used to filter the main features of the image.

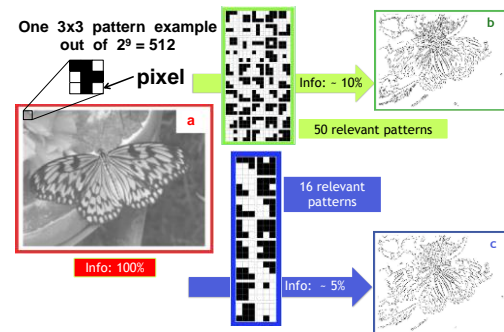


Fig. 1: natural image (a) and corresponding filtered images(b,c)

The pictures on the right (b,c) show the quality of the filtered images. The butterfly can be clearly recognized even if the image information is reduced at the level of 10% or less of the original content. The associative memory works as an edge detector implementation able to extract the salient features.

The pattern is defined as the collection of pixels contained in a 3×3 pixel square, as shown above the butterfly image (a) in Fig. 1. Each square is converted in a 9 bit sequence (each bit is 1 for a black pixel and zero for a white one for B/W) or an 18 bit sequence in case of 4 levels of grey (2 bits/pixel). The bit sequence is used to identify the pattern. Starting from the left top corner the image is scanned by the 3×3 square moved in steps of one pixel toward the right. When the row is finished, the square is moved one pixel down to scan again the raw from the left to the right. Each pattern detected in the figure during the scan is compared to the set of “relevant patterns” predefined by a training phase. It is rejected if it does not match any of them; it goes back in its position in the picture if it is

accepted. Fig. 1 shows two collections of relevant patterns for two different selections. The 16 patterns in the blue box produce a larger image compression than the 50 patterns in the green box. The smaller is the set of chosen patterns the stronger the information reduction that is achieved in the end.

Analyzing images with 4 or 8 levels of grey or using 3-D images increases the number of possible and relevant patterns. The pattern in the 3-D case is not a square, but a cube of pixels: a set of three 3×3 squares taken from 3 subsequent frames. Each pattern for B/W is made of 27 bits corresponding to 2^{27} possible patterns. If 4 levels of grey are used the total number of patterns becomes 2^{54} .

A. Implementation

The algorithm is divided in two main parts: “training” phase and the “Real-Time patterns recognition” phase, what we call the “data taking” phase. Most of the functions are executed by the FPGA with the only exception of the data taking, that is executed by the AM under the FPGA control. We have estimated the processing latency for the data taking exploiting the long AM experience accumulated in FTK [6]. For the training, instead, we implemented the logic on a Xilinx Kintex Ultrascale *XCKU040* of a KCU105 evaluation board, easily connectable to an external PC (or a video camera) and to a set of AM chips [7]. We evaluate the training timing performance directly on the new hardware.

The Training Phase is subdivided in the following steps:

1. *Calculation of the pattern appearance frequencies:* The embedded system receives the image bit-streams (e.g., data from a PC or a video camera). The FPGA partitions/reorganizes the input data into the small 3×3 pixel patterns. Then, for each possible pattern, the FPGA calculates the occurrence frequency in the processed images/frames, using a large set of training images, to measure the frequencies with precision. When the environment and the lighting conditions change, the training has to be repeated in order to identify the relative patterns set suitable for the new environment. Therefore a continuous real-time training execution is required to allow the device adapt itself autonomously to the different conditions of the images that it observes.
2. *Pattern selection:* the system must decide which set of patterns is “relevant”, to be selected for memory storage and later use. We adopt the hypothesis described in [5] to maximize the capability to recognize shapes, i.e., maximum entropy is a measure of optimization. The set of patterns that produces the largest amount of entropy allowed by system limitations (size of the memory to store patterns and output bandwidth) is the best set of relevant patterns. In [5] are described the details of the selection. The selected patterns have to be written inside the AM bank for the following data taking phase.

We implemented the training for 2D B/W images (Fig. 2).

The *FPGA* needs to perform training in real-time for demanding streaming video applications. Several optimization techniques are used to achieve the best performance possible

in the hardware implementation. The video frames are stored in the external memory before being transferred in an internal frame buffer. As soon as enough data has been transferred for the 3×3 patterns to be formed, a pattern identification matrix begins to be loaded. It identifies and propagates two patterns per clock cycle to the pattern accumulators. The accumulators are specifically designed to facilitate successive accumulation in the same memory location (“fall through” data logic). As soon as the whole image sample has been read, the pattern frequency is calculated by taking advantage the *FPGA* DSP slices. The architecture is generic and parametric to allow easier adaptation for the implementations of the more complex 3-D and 4 levels of grey cases.

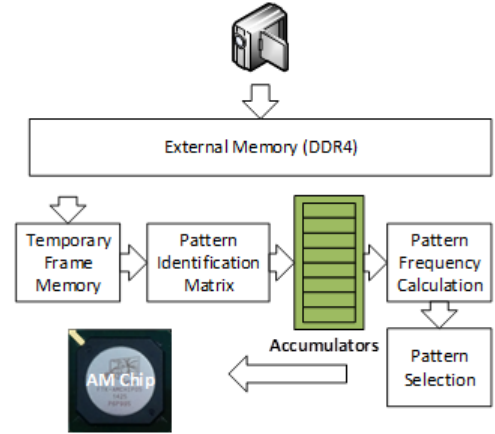


Fig. 2: Training Phase Block Diagram

III. PRESENTED DEMO

We will present the hardware infrastructure of the system and the simulation framework. We will demonstrate by the simulation framework of the system the potential of the brain emulation algorithm for use in generic image processing as well as biomedical applications (e.g. MRI image processing). The demo will present the impact of the design parameters on the system output (number of selected patterns, selection of B/W or grey scale images etc.).

REFERENCES

- [1] ATLAS Collaboration, “The Fast Tracker (FTK) Technical Design Report” CERN-LHCC-2013-007 ; ATLAS-TDR-021; available online: <https://cds.cern.ch/record/1552953>
- [2] ATLAS Collaboration Phys.Lett. B716, 1 (2012);
- [3] M. Dell’Orso and L. Ristori, “VLSI Structures Track Finding”, *Nucl. Instr. and Meth. A*, vol. 278, pp. 436-440, 1989.
- [4] W. Smith, “Triggering at LHC Experiments”, *Nucl. Instr. and Meth. A*, vol. 478, pp. 62-67, 2002.
- [5] M. Del Viva, G. Punzi, and D. Benedetti. Information and perception of meaningful patterns. *PloS one* 8.7 (2013): e69154
- [6] S. Citraro et al., “Highly Parallelized Pattern Matching Hardware for Fast Tracking at Hadron Colliders”, submitted to IEEE TNS.
- [7] A. Andreani, et al., “Characterisation of an Associative Memory Chip for high-energy physics experiments,” in *Proc. I2MTC*, 2014, Montevideo. pp. 1487 – 1