OPAC

Optimization of Pattern Matching Compression Algorithms for GPUs

The OPAC project studied the optimization of high performance image coding algorithms, for many-core (GPUs) and multicore platforms (CPUs), relying on several parallel standards such as openMP and openCL. A broad range of platforms were studied, namely, high end servers, laptops, embedded systems and mobile devices from both the raw performance and energy consumption point of view.

Main Project Team	
Patrício R. Domingues	**MSP-Lr**
Nuno M. M. Rodrigues	MSP-Lr
Sérgio M. M. Faria	MSP-Lr
Gabriel Falcão	MSP-Co
Pedro M.M. Pereira	MSP-Lr
João Silva	ESTG /PI-Lr
Pedro Cordeiro	FCTUC

Indicators	
Funding	32k €
Journal Papers	1
Conference Papers	3
Concluded MSc:	3

Two Main Publications

Pedro M. M. Pereira, P. Domingues, Nuno M. M. Rodrigues, G. Falcão, S.M.M. Faria, Optimizing GPU code for CPU execution using OpenCL and vectorization: a case study on image coding, ICA3PP: "16th International Conference on Algorithms and Architectures for Parallel Processing," Granada, Spain, Vol. 1, pp. 1 - 8, December, 2016.

Pedro M. M. Pereira, P. Domingues, Nuno M. M. Rodrigues, G. Falcão, S.M.M. Faria, Assessing the Performance and Energy Usage of Multi-CPUs, Multi-Core and Many-Core Systems : The MMP Image Encoder Case Study, "International Journal of Distributed and Parallel systems (IJDPS)", Vol. 7, No. 5, pp. 1 - 20, September, 2016

PROJECT WEBPAGE URL
http://bit.ly/OPAC-IT

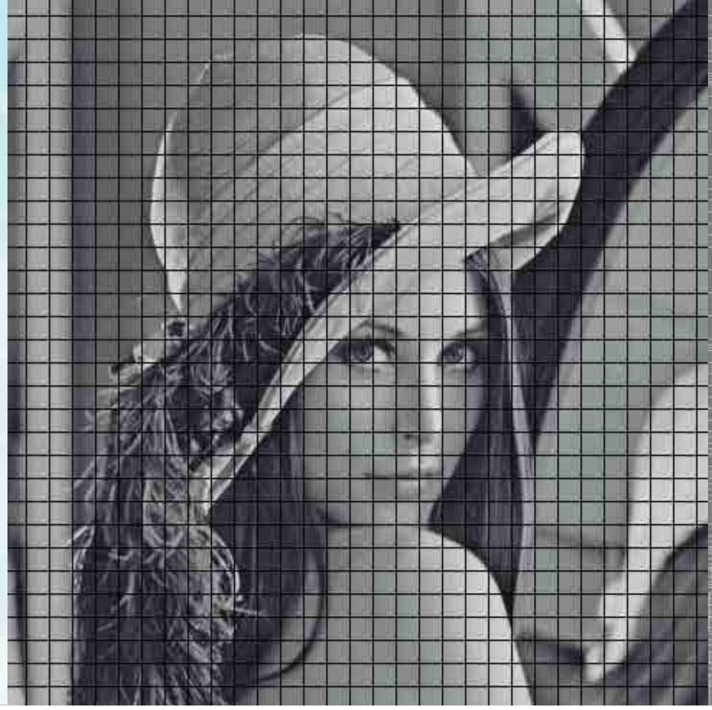

Fig. 1 MMP's Input Blocks on Lenna's 512x512 pixels.

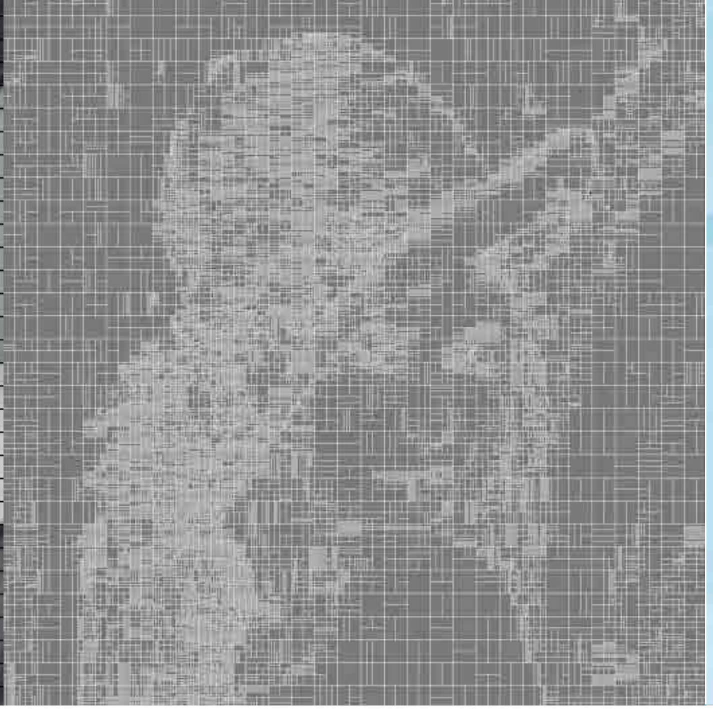

Fig. 2 Explicit pattern matching for MMP's encoding of Lenna 512x512 pixels..

GENERAL MOTIVATION AND OBJECTIVES

Compression of digital images and videos is mandatory to contain the data deluge and to preserve network bandwidth. However, modern image/video compression algorithms achieves high compression ratio at the cost of computational complexity, requiring massive amounts of computational power and energy, and long encoding time. This is also the case for the Multimedia Multiscale Pattern algorithm, which has been developed by the members of the IT research team in the past decade. One of the main goal of the OPAC project was to optimize the raw performance of the algorithms for a broad range of platforms – high end servers, laptops, embedded systems and mobile devices –, while assessing the energy consumption. OPAC also focused on implementing CPU optimized versions, resorting to the OpenMP and OpenCL standards, creating CPU/multicore versions that can be fairly compared to GPU/manycores one.

CHALLENGE

OPAC tackled several challenges. First, pattern-based compression algorithms have limited parallelism, since the compression of an input block is linked to the outcome of the previous block and so on. Therefore, opportunities for parallelism need to be explored in the inner-block compression. Second, although parallel standards allow for code portability across platforms, performance is not portable. For instance, an OpenCL code tuned for CPUs can be run effortlessly on GPUs, but it will not deliver the top performance that can be achieved with a GPU-optimized OpenCL code.

First challenge was addressed by optimizing well known algorithms, used by image/video encoding. This was the case for the Least Square Prediction (LSP) and the Walsh Hadamard transform. The second set of challenges was tackled by building many fine-tuned versions of the algorithms for all the available platforms.

WORK DESCRIPTION AND ACHIEVEMENTS

OpenMP and OpenCL versions of MMP tuned for CPU were developed. In the case of the OpenCL standards, existing GPU-optimized versions were adapted and later optimized for multicore/multiple CPU machines. Highly optimized versions of OpenMP were built, leveraging the parallelism available within the intra-block compression. All of the versions were assessed for raw performance and energy consumption on a wide range of hardware platforms, namely high end multiple CPUs servers, Laptops, and System on a Chip (Raspberry Pi 2 and NVIDIA's Jetson TK1). Experience and knowledge were learnt in collecting and processing simultaneously performance and energy consumption measurements. Another important achievement was the use of the challenging SIMD instructions to optimize CPU-versions of image encoding software. This was done within the OpenCL platforms, boosting the performance of CPU-based code, while maintaining code portability.

Project OPAC also contributed to the optimization of an HEVC-based geometric transformation module for holographic images, applying optimization techniques to reduce the execution length from 7 days to less than a day. During this task, the team identified the opportunity to optimize the widely used Walsh-Hadamard transform (WHT), providing a tuned OpenCL implementation of WHT. The resulting optimized version is available as open source on github.

LSP is another individual algorithm tackled by the OPAC project. LSP is important across a wide area of signal processing, being used in image/video compression for prediction. In OPAC, an OpenCL version of LSP targeting GPU, was tuned and assessed, both for raw performance and energy consumption on a mobile Snapdragon 805 board. The final code yields fast execution times, while consuming a fraction of the energy of other platforms.

In conclusion, the OPAC project achieved several meaningful results: use of OpenMP and OpenCL for optimizing CPU code; Use of CPU's SIMD instructions within the context of OpenCL; study and OpenCL's Optimized version of the LSP and the Walsh-Hadamard algorithms; coupling of energy consumption measurements with execution time performance measurements.