

Editorial

VLSI Circuits, Systems, and Architectures for Advanced Image and Video Compression Standards

Maurizio Martina,¹ Muhammad Shafique,² and Andrey Norkin³

¹ VLSI Lab, Department of Electronics and Telecommunications, Polytechnic University of Turin, Corso Duca degli Abruzzi 24, I-10129 Torino, Italy

² Chair for Embedded Systems (CES), Department of Computer Science, Karlsruhe Institute of Technology (KIT) Kaiserstrasse 12, 76131 Karlsruhe, Germany

³ Ericsson Research, Torshamnsgatan 23, 164 83 Stockholm, Sweden

Correspondence should be addressed to Maurizio Martina, maurizio.martina@polito.it

Received 26 August 2012; Accepted 26 August 2012

Copyright © 2012 Maurizio Martina et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

In the last ten years image and video compression research developed new techniques and tools to achieve high compression ratios even when high quality is required. These techniques involve several parts of the compression/decompression chain including transform stages, entropy coding, motion estimation, intraprediction, and filtering.

At the same time image and video compression standards were ready to identify the most relevant novelties and incorporate them. In this scenario JPEG2000 and MPEG4-part 10, also known as H.264/advanced video coding (AVC), are two important examples of standards able to perform significantly better than their ancestors JPEG and MPEG2. Recently, stemming from the AVC standard, video compression started facing new challenges. The increasing request for high quality, virtual reality, and immersive gaming highlighted how video compression standards have to face not only high definition and high rate sequences but also 3D or multiview systems. These challenges are part of the high efficiency video coding (HEVC) and multiview video coding (MVC) standards. Unfortunately, both HEVC and MVC standards yield to high complexity burden.

In particular, even if some works on the implementation of the MVC standard have already been proposed in the literature, several critical aspects, including performance, memory management, power consumption, and reconfigurability are still open issues. On the other hand, HEVC is the ultimate standard for video compression issued by the Joint

Collaborative Team on Video Coding. As a consequence, works available in the literature mainly address new tools and performance of the HEVC standard. Thus, results on complexity and implementation issues are still on going. In this scenario, advances in camera and display technologies and continuously increasing users' sensation for true three-dimensional (3D) reality have led to the evolution of new 3D video services with multiview videos, like true-3DTV, free viewpoint TV (FTV), realistic-TV, in-car 3D-infotainment, 3D-personal video recording and playback, 3D-surveillance, immersive video conferencing, high-end medical imaging, remote eHealth applications like 3D-teleoperation theaters, telework office, teleshopping, and so forth. The feasibility of multiview video recording of 2–4 views on mobile devices has been demonstrated by the early case studies on 3D-camcorders/3D-mobile phones by Panasonic, Sharp Lynx, Fujifilm, Sony, Minox, etc. However, multiview video processing of 4–8 views for mobile devices and of 16–>100 views for advanced 3D applications (3DTV/FTV, 3D-medical imaging, 3D surveillance, etc.) is foreseen to be adopted in 2015–2020 at a wide range.

On the other hand, HEVC [4] is a next generation video standard, which is regarded as a next major step in video coding standardization after AVC. The standardization has been conducted jointly by VCEG and MPEG in a Joint Collaborative Team on Video Coding (JCT-VC). The standardization has reached a committee draft stage in February 2012. When starting the standardization, the goal was to achieve twice better compression efficiency compared to

AVC standard. That means that the same subjective picture quality should be achieved at twice lower bitrate compared to High Profile of AVC. The standard is primarily focused on coding higher-resolution video, such as HD and beyond. Higher-resolution videos often have different signal statistics from the lower-resolution content. Moreover, coding higher-resolution video also generates higher requirements to the compression ratio as well as to computational complexity of both encoding and decoding. Higher resolutions and higher frame rates set restriction on standard's computational complexity, which was taken into account when working on the standard. Another aspect of video compression that was addressed in developing HEVC is a requirement of easy parallelization, which facilitates higher resolution coding by performing encoding and decoding of video in parallel on multiple processors.

In order to meet up the present and future demands of different multimedia applications, one interesting research direction concerns the development of unified video decoders that can support all popular video standards on a single platform. It is worth noting that several video compression standards use the DCT-based image compression. Since the DCT computation and quantization process are computation intensive, several algorithms are proposed in literature for computing them efficiently in dedicated hardware. Research in this domain can be classified into three parts: (i) reduction of the number of arithmetic operators required for DCT computation, (ii) computation of DCT using multiple constant multiplication schemes, (iii) optimization of the DCT computation in the context of image and video encoding.

Another important aspect related to the complexity of video-coding standards is the large amount of coding tools used to achieve high compression rates while preserving the video visual quality. As an example, the prediction steps of H.264/AVC (composed by intraframe prediction and interframes prediction) are responsible for the main contribution of compression provided by this standard, which results in about 50% fewer bits to represent a video when compared to MPEG-2. This result is achieved through the insertion of a large number of coding modes in the prediction steps and selecting the best one to encode each macroblock (MB). However, the computation of a large number of prediction modes provided by H.264/AVC standard is extremely expensive in terms of computational complexity. In particular for HD formats several millions of complete encoding operations are needed for each video frame. In order to support real time video encoding and decoding, specific architectures are developed. As an example, multicore architectures have the potential to meet the performance levels required by the realtime coding of HD video resolutions. However, to exploit multicore architectures, several problems have to be faced, such as the subdivision of an encoder application in modules that can be executed in parallel. Once a good partitioning is achieved, the optimization of a video encoder should take advantage of the data level parallelism to increase the performance of each encoder module running on the processing element of the architecture. A common approach is to use single instruction multiple data instructions to

exploit the data level parallelism during the execution. However, several points have not been addressed yet, for example, how the data level parallelism is exploited by SIMD and which instructions are more useful in video processing. Moreover, different instruction set architectures (ISAs) are available in modern processors and comparing them is an important step toward efficient implementation of video encoders on SIMD architectures.

In 2007, the notion of electronic system level design (ESLD) has been introduced as a solution to decrease the time to market using high-level synthesis. In this context, CAL was introduced as a general-use data flow target agnostic language based on the data flow process network model of computation. The MPEG community standardized the RVC-CAL language in the reconfigurable video coding (RVC) standard. This standard provides a framework to describe the different functions of a codec as a network of functional blocks developed in RVC-CAL and called actors. Some hardware compilers of RVC-CAL were developed, but their limitation is the fact that they cannot compile high-level structures of the language so these structures have to be manually transformed. Thus, this research field requires further investigation. This special issue is dedicated to research problems and innovative solutions introduced above in all aspects of design and architecture addressing realization issues of cutting-edge standards for image and video compression. The authors have focused on different aspects including (i) VLSI architectures for computationally intensive blocks, such as the DCT and the intraframe coding mode, (ii) automatic code generation and multicore implementation of complex MPEG4 and H.264 video encoders. Due to the increasing importance of stereo and 3D video processing an invited paper dealing with this topic is included in the issue.

2. VLSI Architectures for Computationally Intensive Blocks

As long as new standards have been developed, several research efforts were spent to design efficient VLSI architectures for computationally intensive blocks. Stemming from the works of Chen and Loeffler, several techniques were proposed to reduce the complexity and to increase the flexibility of architectures for the computation of the DCT. Formal techniques as subexpression elimination and canonical sign digit (CSD) representation are viable techniques to optimize architectures for the computation of the DCT. The paper "optimized architecture using a novel subexpression elimination on loeffler algorithm for DCT based image compression" by M. Jridi et al. presents a novel common sub-expression elimination technique that is used with the canonical sign digit (CSD) representation to reduce the complexity of the Loeffler algorithm for DCT implementation. An FPGA-based implementation is provided with video quality results in terms of peak signal to noise ratio (PSNR). Other approaches are based on the factorizing the DCT by the means of simpler and modular structures. Some these ideas were already described in Rao Kamisetty works, but their impact on VLSI implementation has not been completely studied especially when flexibility and multistandard

requirement have to be taken into account. In the paper “N point DCT VLSI architecture for emerging HEVC standard” by A. Ahmed et al. a variable-sized DCT architecture is presented. The architecture stems from the Walsh-Hadamard transform and uses the lifting scheme to reduce the number of multiplications required. The authors provide hardware implementation results for a 90 nm ASIC technology. The paper “low cost design of a hybrid architecture of integer inverse DCT for H.264, VC-1, AVS, and HEVC” by M. Martuza and K. A. Wahid proposes a unified architecture for computation of the integer inverse DCT of multiple modern video codecs. Moreover, the authors show both FPGA- and ASIC-based implementation results.

It is known that the DCT is only one of the most computationally intensive blocks in video compression systems. Several other blocks, including motion estimation and entropy coding, are known to be a significant part of the computational burden of video compression systems. With the H.264/AVC standard very high quality is obtained even with very low bit rates. Such an impressive improvement is obtained due to the possibility to employ a large number of new tools, as intraprediction, coupled with several coding modes. As a consequence, the optimal choice of a coding mode is a very computationally intensive task. Thus, techniques for the fast identification of the optimal or nearly-optimal coding mode are of paramount importance. The paper “low complexity hierarchical mode decision algorithms targeting VLSI architecture design for the H.264/AVC video encoder” by G. Corr ea et al. presents a set of heuristic algorithms targeting hardware architectures that lead to earlier selection of one encoding mode. The resulting solution leads to a significant complexity reduction in the encoding process at the cost of a relatively small compression performance penalty.

3. Automatic Code Generation and Multicore Implementation

The development of VLSI architectures, circuits, and systems for video compression is a time-consuming task. As a consequence, industries are always working hard to be ready with product that are on the cutting-edge of the available technology. Automatic code generation is a very appealing strategy to speed up the design process and to reduce recurrent costs. In the paper “automatic generation of optimized and synthesizable hardware implementation from high-level dataflow programs” by K. Jerbi et al., the authors describe a methodology that from a high-level language called Cal Actor Language (CAL), which is target agnostic, automatically generates image/video hardware implementations able to largely satisfy the real time constraints for an embedded design on FPGA.

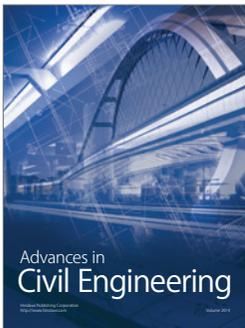
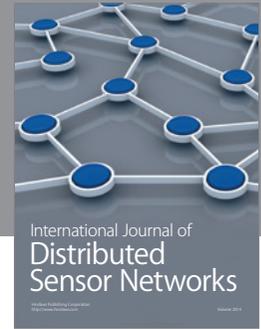
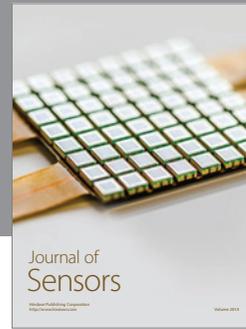
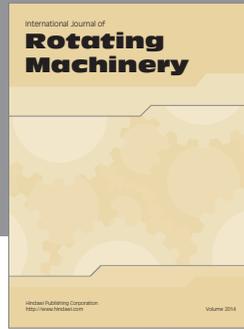
Other directions have been investigated to reduce the time required to develop hardware components. The availability of high performance multicore processors (e.g. GPUs) is pushing several researchers to use software solutions even for very complex systems as video encoders. This direction is investigated in the paper “an efficient multi-core SIMD implementation for H.264/AVC encoder”, by M. Bariani et al.

where the optimization process of a H.264/AVC encoder on three different architectures with emphasis on the use of SIMD extensions is described. Experimental results are given for all the three architectures.

4. Depth Maps Extraction

The availability of devices for 3D displaying together with standards for 3D and multiview video processing has highlighted how 3D video is a very challenging topic. One of the main issues is related to the fact that the signals acquired from cameras are 2D, and signal processing is required to merge data together to create a 3D video sequence or to create a new view of the scene. In particular, it is widely recognized that extracting depth maps from 2D images is one of the key elements to create 3D video. In this perspective, the invited paper “hardware design considerations for edge-accelerated stereo correspondence algorithms” by C. Ttofis and T. Theocharides deals with this hot topic: extracting depth maps from 2D images. In particular, in this work the authors present an overview of the use of edge information as a means to accelerate hardware implementations of stereo correspondence algorithms. Their approach restricts the stereo correspondence algorithm only to the edges of the input images rather than to all image points, thus resulting in a considerable reduction of the search space. The resulting algorithms are suited to achieve real-time processing speed in embedded computer vision systems. For both algorithms, optimized FPGA architectures are presented.

*Maurizio Martina
Muhammad Shafique
Andrey Norkin*



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

