

OPTIMIZATION AND COMPARISON OF COMPUTATIONAL COMPLEXITIES OF STANDARD COMPLIANT VIDEO DECODERS ON SIMD PROCESSOR

Anjaneya Prasad, Aditya Mittal, Mangesh Kumar, Prasad Sankaran, Ruturaj Chandpur

Emuzed India Private Limited, Bangalore, India.

ABSTRACT

The intrinsic nature of multimedia applications has been targeted for exploiting single-instruction, multiple-data (SIMD) extensions to instruction architectures for most of the embedded processors. In particular, SIMD instructions can be most effective in multimedia applications, which have simple operations on multiple and small data types, mostly 8-bit or 16-bit samples. In this paper, optimization and comparison of computational complexities of standard video decoders has been done using the Intel Wireless MMX technology, to show the overall speedup by SIMD style coding.

Index Terms— SIMD, MPEG-4, VC-1, H.264

1. INTRODUCTION

Multimedia devices, like PDA's carrying capabilities of today's most compelling high-end consumer electronics are almost ubiquitous nowadays. It is obvious that more and more new features, new functionalities will get added in near future. This, in turn, will put enormous load on the underlying processor. The amount of processing power available in any mobile device is quite limited and has a direct impact on battery power consumption. Therefore, it is essential to optimize the applications as much as possible.

The video coding standards are being under development for various applications; the purpose includes better picture quality, higher coding efficiency and more error robustness. The various standards MPEG4 [1], VC-1 [2], [3] and H.264 [4], [5] are aimed at catering the needs of wide range of applications like Internet based subscription services and various wired and wireless consumer electronics devices like mobile phones. Algorithms present in these standard compliant decoders require enormous computations. Hence, vigorous optimizations of these algorithms are required to achieve real time decoder solution. On the other hand, SIMD extensions to the embedded processors have come to support the ever-increasing requirements by providing data-level parallelism, which improves the performance for multimedia applications.

The present work is aimed at optimizing decoding functionality of MPEG4 (SP), VC-1 (SP) and H.264 (BP) on SIMD platform and the comparative analysis of the complexities of various modules in these decoders. Intel Bulverde is based on Intel Xscale core architecture and supports SIMD instructions. This processor has been used for optimization and comparative analysis. It is particularly suitable for wireless devices because of its high processing power and very low power consumption.

The remainder of the paper is organized as follows. In section 2, overview of the Bulverde processor is presented, followed by overview of standard video decoders, i.e., MPEG4, VC-1 and H.264, in section 3. Section 4 describes the algorithmic optimizations for MPEG4, VC-1 and H.264. Section 5 describes the memory optimizations for the three decoders. Results are presented in Section 6. Concluding remarks are given in Section 7 and references are listed in Section 8.

2. BULVERDE PROCESSOR

Bulverde [6] is Intel application processor based on Intel XScale core and Wireless Multimedia extension (WMMX). WMMX technology is a high-performance, low-power, seamless extension to Intel XScale micro architecture. This technology offers a powerful set of new instructions that will help in enhancing the multimedia applications on handheld devices. The powerful 64-bit SIMD architecture of WMMX gives a performance enhancement to many applications including multimedia codecs. Fig 1 shows the architectural support for WMMX and XScale core.

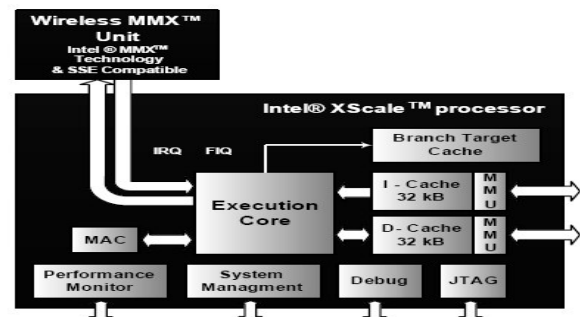


Fig. 1 Architectural support to WMMX extension

WMMX technology defines four data types, packed byte, packed half word, packed word, and double word. WMMX has introduced 16 - 64 bit registers. The instructions can be classified as add, subtract, multiply, compare, and shift. The packed and saturating arithmetic instructions are very useful for video compression applications. By executing same operations upon two, four, or eight data elements at a time, Wireless MMX™ technology speeds up applications that exhibit data level parallelism.

3. STANDARD DECODERS

The standard decoders that are widely used in the real time applications are MPEG-4 simple profile decoder, VC-1 simple profile decoder and H.264 base line profile decoder.

3.1. MPEG-4 Simple Profile (SP) Decoder

MPEG-4 has been introduced to increase error robustness to support wireless networks, and better support for low bit rate applications. MPEG-4 Simple Profile is the simplest profile in MPEG-4 and supports Intra coding with context adaptive intra DCT, AC/DC prediction, and Inter coding with unrestricted motion vectors, variable block size motion compensation, and four motion vectors.

3.2. VC-1 Simple Profile (SP) Decoder

Windows Media formats are the leading formats for audio and video subscription services and streaming media on the Internet. Windows Media Video was originally a Microsoft proprietary algorithm that is now standardized by SMPTE as VC-1. VC-1 simple profile decoder is the simplest profile in VC-1 and supports Intra and Inter coding with bi-linear and bi-cubic motion compensation. It also supports adaptive block size transform and overlapped transform.

3.3. H.264 Baseline Profile (BP) Decoder

H.264/AVC video coding standard has been introduced with significant enhancements in both video coding efficiency and flexibility over a variety of network domains. In video coding layer (VCL), some of important enhancements are the use of a small block-size (4x4) exact-match transform, adaptive in-loop deblocking filter and motion-prediction capability.

H.264 defined three types of profiles. The baseline profile is the simplest and supports intra and inter-coding, and entropy coding with context-adaptive variable-length coding (CAVLC).

In all the above standards Inter prediction, loop filtering, inverse transform, and intra prediction are most time consuming modules and these are implemented using SIMD instructions for speedup.

4. ALGORITHMIC OPTIMIZATIONS

In this section the algorithmic optimizations of various decoders is described.

4.1. Optimizations in MPEG4 SP

Inter prediction and inverse transform are the complex modules, which are optimized using SIMD.

4.1.1. Inter Prediction

Inter prediction entails predicting the Macro Block (MB) or Block from the reference frame by using the motion vectors. Motion compensation can be either at 16x16 level or 8x8 level, and motion vectors can be either integer or half-pel. For the full pel, data will be copied from reference frame and in half pel predicted data is obtained by applying bi-linear interpolation on the reference data.

Copying of data from reference frame exhibits data parallelism, which can be well optimized by copying array of elements from the reference frame. For bi-linear interpolation averaging Instructions, which do both averaging and rounding, can be used to ameliorate performance.

4.1.2. Inverse Transform

The transform used in MPEG4 is 8x8 DCT. Inverse transform is applied on a block of size 8x8. Data parallelism can be obtained by processing four elements; each element is of size 16 bits. WMMX register, which is 64-bit, can be used to hold four elements and can perform DCT operations in parallel. DCT involves multiplication and addition operations, which are optimized using SIMD multiplication instructions. Four multiplications and four additions are performed in a single instruction.

4.2. Optimizations in VC-1 SP

The distribution of time complexity amongst major subsystems was analyzed and Inter prediction, inverse transform are the order of the averaged time consuming modules.

4.2.1. Inter Prediction

Inter-prediction involves filtering with either bi-cubic (4 taps) or bi-linear (2 taps) filter. Pixel values from the reference frame are multiplied with filter coefficients and averaged. Once the filter type is fixed, the operation is same for the entire block and these multiplications and averaging is effectively implemented using SIMD.

4.2.2. Inverse Transform

VC-1 takes the approach of allowing 8x8 blocks to be encoded using either one 8x8, two horizontally stacked 8x4s, two vertically stacked 4x8's or four 4x4 block

transforms. This allows VC-1 to use the transform size and shape that is best suited for the underlying data. Input as well as the intermediate results of inverse transform is of 16-bit precision. Four pixels can be processed with one register of 64-bit size, with the help of SIMD multiplications.

4.3. Optimizations in H.264 BP

The complex modules in H.264 base line profile decoder are Inter prediction, Deblocking, Inverse transform and Intra prediction. These modules are optimized using SIMD.

4.3.1. Inter prediction

Inter prediction in H.264 involves filtering of pixels using 6 tap filter and these operations can be applied in parallel by the efficient loading of registers which reduces the MIPS required for the module drastically.

4.3.2. Deblocking

Deblocking in H.264 is very complex and takes nearly 1/3rd of the total decoding time. Most of the 4x4 block edges will be filtered. This filtering involves computing of absolute differences between the elements on either side of the edge and computing the value of filtered pixels. These operations are implemented using SIMD, which improves the performance. Filtering of vertical edges in a macroblock is performed in the horizontal direction in order to reduce the number of memory loads. By doing filtering in horizontal direction the processed pixels for one vertical edge can be reused for processing the next vertical edge. This reduced the number of memory loads required for vertical filtering by 50%.

4.3.3. Inverse Transform

Inverse transform in H.264 is integer transform and it involves only additions and shifts. Inverse transform is applied on a block of size 4x4. Each element in 4x4 block is 16 bits in length. So to hold an array of 4 elements, 64-bit register is required and WMMX registers are well suited for this. Inverse transform operations are applied on arrays instead of elements, to improve the performance.

4.3.4. Intra prediction

4x4 and 16x16 horizontal prediction, vertical prediction and DC prediction in Intra prediction involves copying of data from top or left block as indicated by the prediction mode. This copying is effective when multiple elements are loaded or stored in a single operation, which will improve the data through put.

4.3.5. Reconstruction

Reconstruction involves addition of 8-bit predicted data with the 16-bit inverse transform output and saturating the result to 8-bits. This module is present in all the decoders and it is

done for all the Inter macro-blocks. Reconstruction is optimized using parallel addition and packing instructions.

5. MEMORY OPTIMIZATIONS

The memory optimizations are cache related optimizations. Bulverde processor has 32KBytes data cache and 32 Kbytes instruction cache. Cache memory works on the principle of locality of reference. As long as the data access is sequential cache works well. When the data access is non-sequential, data has to be loaded from main memory to the cache memory. This loading of data takes tens of cycles because of speed mismatch between the processor and the main memory. This latency can be avoided by efficiently preloading the required data from the main memory to the cache memory. While preloading, optimal preload distance is maintained in order to reduce the cache misses. This preloading is effectively implemented in all the complex modules. Inter prediction can access non-sequential data from the reference frame depending on the motion vector. This data is preloaded before the actual computations in order to reduce the cache miss and subsequently improving the performance.

6. RESULTS

This section illustrates the performance improvement as a result of optimizations for the three decoders. The complexity ratio of VC-1 and H.264 with respect to MPEG-4 is computed. Also herein, percentage complexity of the different modules of these decoders is compared. Performance improvement in percentage on Bulverde is given in [Table 1](#). Complexity ratios are given in [Table 2](#) and percentage complexities of the various modules in the decoders are compared in [Table 3](#). Test streams are generated after enabling all the complex modules in the respective encoders. Streams used to compare the performance are foreman and football streams. The percentage increase in number of frames decoded per second on Intel Main Stone board running at 520 MHz is taken as the performance metric in [Table 1](#). The figures under the respective decoders give the performance improvement after Bulverde assembly optimization with respect to optimized C code. The compiler used does not employ any SIMD optimization.

Performance improvement (in percentage)			
Stream	MPEG-4	VC-1	H.264
Foreman, QVGA	75	73	94
Foreman, CIF	73	71	92
Football, QVGA	78	67	95
Football, CIF	76	64	100

Table 1: Performance improvement in percentage

Using SIMD architecture, the improvement in H.264 decoder performance is nearly 100% with respect to the optimized C code. Improvements in MPEG-4 and VC-1 are 75% and 70% respectively. Better improvement in H.264 is because data parallelism is more in H.264 over MPEG-4 and VC-1.

The number of frames decoded per second (fps) of VC-1 and H.264 decoder with respect to MPEG-4 fps is taken as performance metric to compute the complexity ratios of the decoders with respect to MPEG4 in [Table 2](#).

Complexity ratio w.r.to MPEG-4			
Stream	MPEG-4	VC-1	H.264
Foreman, QVGA	1	1.45	2
Foreman, CIF	1	1.41	1.96
Football, QVGA	1	1.5	2.1
Football, CIF	1	1.55	2.15

Table 2: Complexity ratio w.r.to MPEG-4

From the above table, it is lucid that H.264 is twice more complex than MPEG-4. This is because loop filtering and intra prediction is not present in MPEG-4. Inter prediction takes only $\frac{1}{2}$ pel motion vector, bi-linear motion compensation in MPEG-4, where as in H.264 it is $\frac{1}{4}$ pel motion vector, 6-tap filter.

VC-1 is approximately 1.5 times more complex than MPEG-4. This is because Overlap transform is not present in MPEG-4. Inter prediction in VC-1 is $\frac{1}{4}$ pel BiCubic, 4-tap or bi-linear filter which is more complex than MPEG-4.

It is also found that H.264 is around 1.5 times more complex than VC-1. This is because loop filtering is not present in VC-1. Inter prediction in H.264 is more complex than VC-1 as the filter length in H.264 is more than in VC-1.

Modules	MPEG-4	VC-1	H.264
Deblocking	--	--	30.1
Overlap Transform	--	2.1	--
Inter Prediction, Reconstruction	43.5	57.7	44.3
Intra Prediction, Reconstruction	1.5	2.5	1.6
Inverse Transform	6.4	9.1	1.7
VLD Decoding Control Flow	48.6	28.6	22.3

(-- Represents not applicable)

Table3: percentage complexities of different modules

[Table 3](#) shows the percentage complexities of different modules in the three decoders. Deblocking in H.264 takes 30% of the total decoding time, and this is the main reason why H.264 is more complex than MPEG-4 and VC-1. Inter prediction in all the three decoders takes around 50% of the total decoding time, as most of the frames are Inter frames in the compressed streams.

7. CONCLUSIONS

This paper outlines the features in MPEG-4 SP, VC-1 SP and H.264 BP and the implementation of these decoders on SIMD platform. Significant improvement in performance is achieved by optimizing the decoders on Bulverde platform. Also mentioned is the performance and comparative analysis for the three decoders. The optimizations done are not specific for Bulverde and can be extended to any SIMD processor.

8. REFERENCES

- [1]. "Information Technology – Generic Coding for Audio-Visual Objects – Part 2: Visual," *MPEG-4 Standard, 14496-2, ISO/IEC Standard*.
- [2]. "Proposed SMPTE Standard for Television: VC-1 Compressed Video Bit stream Format and Decoding Process", SMPTE 421M, 2005-08-23.
- [3]. S. Srinivasan, P. Hsu, T. Holcomb, K. Mukerjee, S. Regunathan, B. Lin, J. Liang, M. Lee and J. Ribas-Corbera, "Windows Media Video 9: overview and applications", *Signal Processing: Image Communication*, Vol.19, No. 9, pp. 851-875, Oct. 2004.
- [4]. Text of ISO/IEC 14496 10 Advanced Video Coding 3rd Edition (ISO/IEC JTC 1/SC 29/WG 11 N6540).
- [5]. I. E. G Richardson, "H.264 and MPEG-4 Video Compression: Video Coding for Next-generation Multimedia", *Wiley*, 2003.
- [6]. Intel® Wireless MMX™ Technology Developer Guide.