The Case for Software Solutions for Real Time Video Applications

igolgi Video Compression Software Technology

Summary

igolgi high quality H.264 video encoder solutions provide exceptional compression performance while providing fast, real time performance on multi-core platforms. This white paper summarizes some of the high quality algorithmic features of the igolgi H.264 encoder and demonstrates the performance gains achieved. For typical video sequences, the igolgi dual pass, High Profile encoder provides 25% improvement over H.264 single pass main profile video encoders.

Introduction

igolgi has developed high quality H.264 and MPEG2 codec solutions targeted for Multi-Core CPUs. This solution was developed with 3 key product requirements in mind:

1. High Video Quality
    a. Utilizing the most advanced tools of the H.264 and MPEG2 standard
    b. Utilizing igolgi novel video encoder algorithms
2. High Scalability on multi-core platforms
    a. Ground up software architecture for multi-core
    b. Ground up parallel video compression algorithms targeted to scale with multi-core
3. High Application Flexibility
    a. A high level of pre-processing, pre-analysis, and codec parameter control
    b. Clean software architecture with plug and play APIs

The igolgi codec is a key component that serves a variety of video applications such as satellite, cable, or IPTV video transmission, internet video distribution, mobile video, and transcoding applications including on Cloud platforms.

Today, many of these applications utilize expensive, inflexible hardware based platforms. igolgi believes general purpose; dense, multi-core CPUs will replace these hardware solutions in the coming years.

Multi-Core Background

The CPU technology trend is strongly moving towards multi-core architectures. This is true for practically all general purpose CPU applications such as PCs, Graphics, and Mobile devices. Processing Power continues to progress at a Moore's law pace in part due to the advances in multi-core density and architecture.

Today, quad-core and hexa-core x86 CPUs are in the market, and in 2010 advances in 6-core and 8-core chips are scheduled to appear. Intel and AMD both have released x86 multi-core architectures providing a wide variety of market choices. Server platforms with 1, 2, or 4 CPU sockets is common, providing very cost effective server platforms with 6, 12, and 24 cores today.

Mobile chips are also moving into multi-core direction with 2-core chips coming to market now, and quad core expected by 2011-2012.

For dedicated applications like video processing and compression, it is a challenge to create native parallel algorithms and associated software architecture that achieves high video quality *and* scalability with the number of cores. This is the challenge that igolgi took on in January 2007 with the result being excellent H.264 and MPEG2 solutions.

Amdahl's law from computer science states that as the number of parallel cores increases, the more parallel the software architecture must be. In Figure 1 below, as the number of cores increases, the impact of non-parallel algorithms grows exponentially. As the CPU market evolves from 10s of cores in 2010 to hundreds of cores in the next 5 years, parallel algorithms will be key for providing efficient solutions.

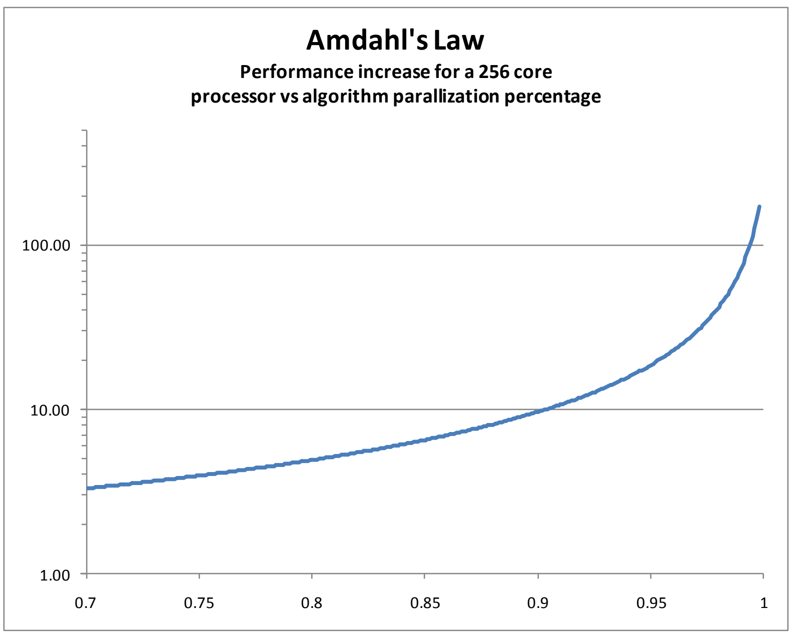

Figure 1: Amdahl's Law for Parallel Computing

At the same time as CPU technology goes multi-core, video technology complexity continues to increase. Video compression complexity as measured as codec complexity and pixel rate, increases every few years. Examples such as the movement from SD to HD (a 6X increase), MPEG2 to H.264 (a 20X or more increase), and future advances such as 3D, 1080p and MPEG-SVC, demonstrate the effect and trends.

Figure 2 shows the trend in multi-core CPU processing power and video encoder compression complexity. As can be seen in the chart, there is a cross-over occurring right now, where multi-core CPU processing power has finally enabled the development of software based, real time, high quality video encoders. igolgi believes this cross-over will be seen as a revolution for traditional video equipment purchasers, who will finally have flexible, off the shelf platforms to run their networks.

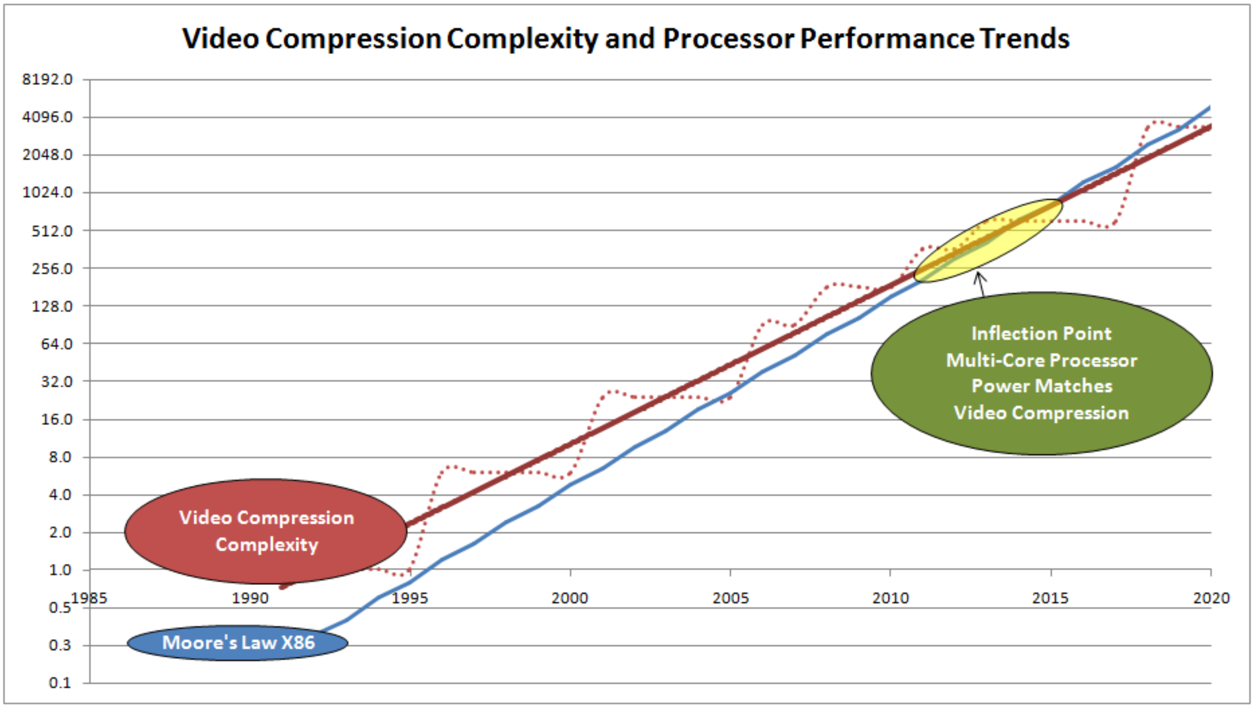

Figure 2 : Inflection Point for Software Platforms

Total Cost of Ownership

The total cost of ownership (TCO) for video distributors is a function of several variables:

- Capital Cost of Equipment
- Maintenance
- Facility Space
- Power and Cooling
- Redundancy
- Cabling

Today, typical solutions for a video workflow are based on cabling together individual hardware based rack mount components. Common video components include video interface and conversion, de-mux, video format conversion, video filtering, decoder (for transcoder applications), encoder, mux, and audio workflow as well. What typically results for common installations is multiple components racked and cabled together one after the other. For redundancy (especially for real time applications), then the complete equipment workflow is repeated. All of these components repeat common elements such as power supplies and management interfaces. Also, typically it is not possible to get exactly the functionality needed in each component, so many times components are used with features that are never used for a specific application, wasting power and cost.

Software based solutions can offer a fully integrated solution and can provide lowest TCO in a number of ways.

- Capital Cost of Equipment
  - Off the shelf IT solutions offer significant cost savings compared to hardware based solutions due to the large volumes of the IT industry. Single computer or Blade based form factors are available that are suited for different installation requirements.
- Maintenance
  - Maintaining a single software platform offers the convenience of a single management interface to control, and limits the box to box cabling and interface incompatibility issues.
- Facility Space
  - For dense installations, Blade based software platforms offer the highest channel density, and provide many economies due to shared(and redundant) power supplies, management interfaces, etc.
- Power and Cooling
  - For a single function, a hardware functional block will usually be lower power than a software functional block. However, for an integrated system, a software solution is lower power due to the benefit of combining functionality onto single CPUs. Also, hardware based solutions do not allow selection of using only the functions a workflow needs, therefore powering unused functions. Software solutions only spend power on the processing that the workflow needs, lowering power.
  - By integrating functions into 1 sever, instead of several hardware rack units stacked together with power hungry video I/O, the total system power is reduced as well. Reducing the number of power supplies (and the AC/DC loss) helps to save power. With blade based solutions especially, power and cooling are extremely well designed and offer highest efficiency.
- Redundancy
  - With integrated solutions onto one software platform, there are less hardware components to repeat for redundancy. In addition, with Blade based form factors, an N+1 redundancy is possible compared to N+N for individual boxes. A single redundant blade in a chassis, can offer full redundancy for any of the other blades. Redundant power supplies and networking are also built into blade chassis.
- Cabling
  - Cabling is often overlooked when considering cost and complexity, but can be significant. Integrated functions onto a software platform can reduce cabling needs, and again, in the case of Blade based computers the cabling can often be integrated into the chassis reducing cabling needs tremendously.
- Flexibility and Upgradability
  - Software solutions allow the most amount of customization for specific applications. Also, as the CPU industry continues to invest and march along at a Moore's law pace, a software solution can be upgraded onto new platforms offering lower cost, higher density, and lower power over time. In a blade based system this is as simple as swapping out blades with no change to cabling or management interfaces.

igolgi H.264 Encoder Overview

The igolgi H.264 encoder high level architecture is shown in Figure 3. It includes the following main features

1. Two-Pass Architecture
    a. This provides highest level of rate control optimization which greatly improves video quality
    b. Flexible latency control all the way to single pass operation
2. Pre-processing functions
    a. Optional components that can provide downscaling, noise filtering or other functions
3. Pre-Analysis functions
    a. Novel image segmentation and analysis technology that isolate textures, edges, and other image features that help drive mode and motion estimation decisions improving quality
4. H.264 core
    a. The core encoder is composed of the key elements of the H.264 work flow such as motion estimation, mode decision, entropy, etc.
5. All algorithms in the design are natively dense multi-core scalable

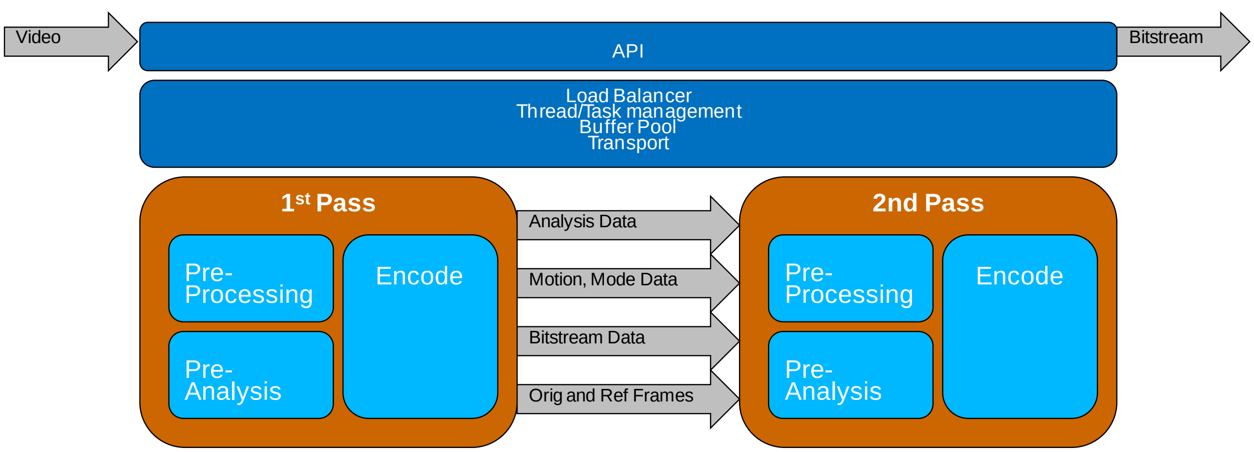

Figure 3 : igolgi high level Video Encoder Architecture

igolgi H.264 Encoder Multi-Core Scalable Software Architecture

The igolgi encoder algorithms and software architecture are parallel in nature. There are 2 layers of parallelism:

1. Macroblock Level Algorithms => Each algorithm was designed to be scalable across N compute cores.
2. Data Level => Data structures are parallel organized to support CPU SIMD instructions

These 2 levels of parallelism allow the igolgi codec to scale (up or down) with the number of cores available. Figure 4 demonstrates the high level architecture. Each Task is allocated on a variable number of video blocks and can share data with neighboring tasks as needed. The diagram demonstrates the seamless scalability in the case of moving from 16 cores to 32.

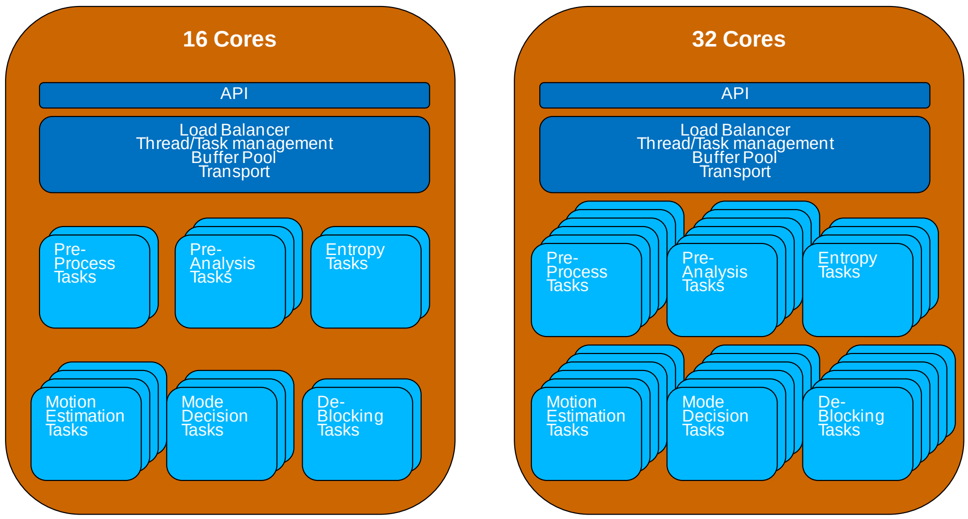

Figure 4 : igolgi parallel architecture and scalability

igolgi H.264 Encoder Video Quality Features

The following sections describe just some of the key elements of the igolgi encoder that enable very high video quality. The igolgi H.264 encoder starts with high quality algorithms in the core such as all block modes, Rate Distortion based mode decision, and intelligent and aggressive motion estimation. This is just the foundation for a good quality encoder. To achieve very high quality, other advances were developed such as:

- Two Pass Rate Control
- Pre-analysis and image segmentation
- High Profile
- Hierarchical GOP structure
- Weighted Prediction

These are just some of the features that enable igolgi high quality encoding which will be covered in this document.

Two Pass Rate Control

Two pass rate control provides a high level of video quality improvement, as well as allowing more accurate bit rate and buffer level management, which is needed in many modern video applications.

There are several aspects to the video quality improvement provided by two-pass rate control schemes:

- Smooth Bit Rate Allocation. Single pass rate control typically results in jumpy quantization allocation in an effort to adjust to the complexity of the video as it is encoded. Two pass rate control smoothes out this variation to optimize the quality over a number of frames, reducing artifacts.
- Improved Mode/Quant Decision: First pass encoding generates a large amount of video characteristic and macroblock level data that is utilized in the second pass mode decision and quantization control algorithms.

Figure 5 demonstrates the difference in bit rate for single pass and dual pass. The dual pass algorithm provides smoother bit rate control and a more constant video quality profile. Typical encoder performance gain for a two-pass rate control scheme is 5-10% compared to single pass. Higher objective gains are common during certain video segments. Additionally, subjective quality improvement can be much higher on video segments due to the smoother PSNR(video quality).

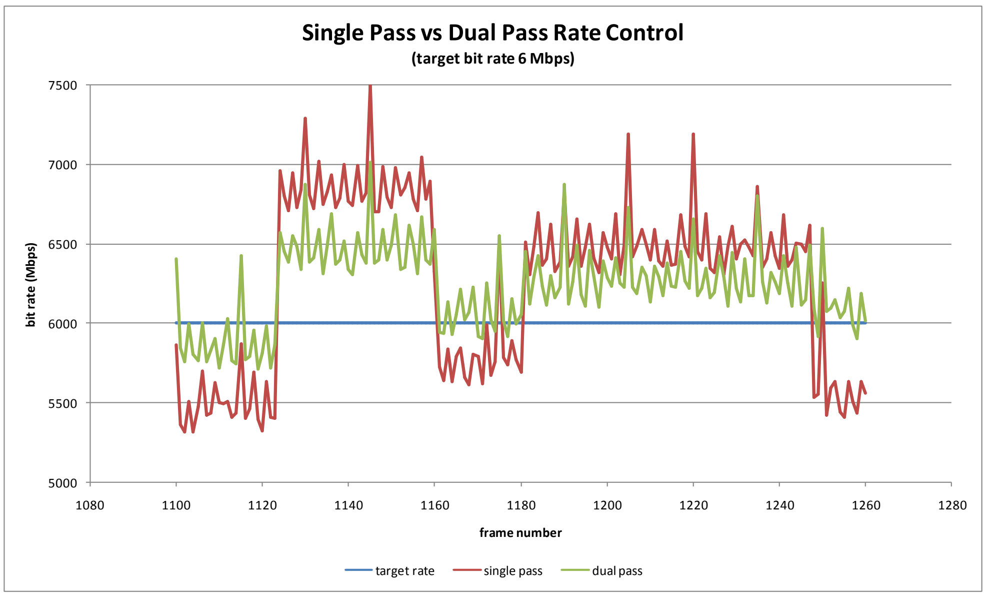

Figure 5 : Benefit of Smoother Video Bit Rate with Dual Pass Encoder

Pre-Analysis

Video content varies widely in texture, motion, color, brightness, sharpness and other characteristics. Video compression encoding is an optimization problem to find the best combination of compression tools (mode, motion, block size, quantization, etc.) to get the most efficient bit stream representation. A brute force approach is to try all combinations of compression tools and choose the best result. This however is impractical for majority of applications, especially real time applications.

igolgi developed novel pre-analysis tools that provide image segmentation and analysis. These tools provide valuable macroblock level video content characteristics that can be leveraged by the core H.264 algorithms. In the example of Figure 6, a single frame of video is analyzed for different textures, edge and skin features. The igolgi core encoder algorithm uses these and other analysis outputs to intelligently optimize the video encoder algorithms for high compression quality improvement.

There are 2 main benefits. First, higher quality video at lower bit rates result since compression decisions are tuned for the content. Second, faster mode decision and other algorithms are created making high quality, real time encoding possible.

Objective measures show typically 5% improvement with pre-analysis based video algorithms. Subjective measures are higher. For example, when skin is detected, then more bits are given to the skin areas resulting in better quality for faces. This results in better human visual subjective quality, as the human eye is more sensitive to artifacts in human faces, than in non-human objects.

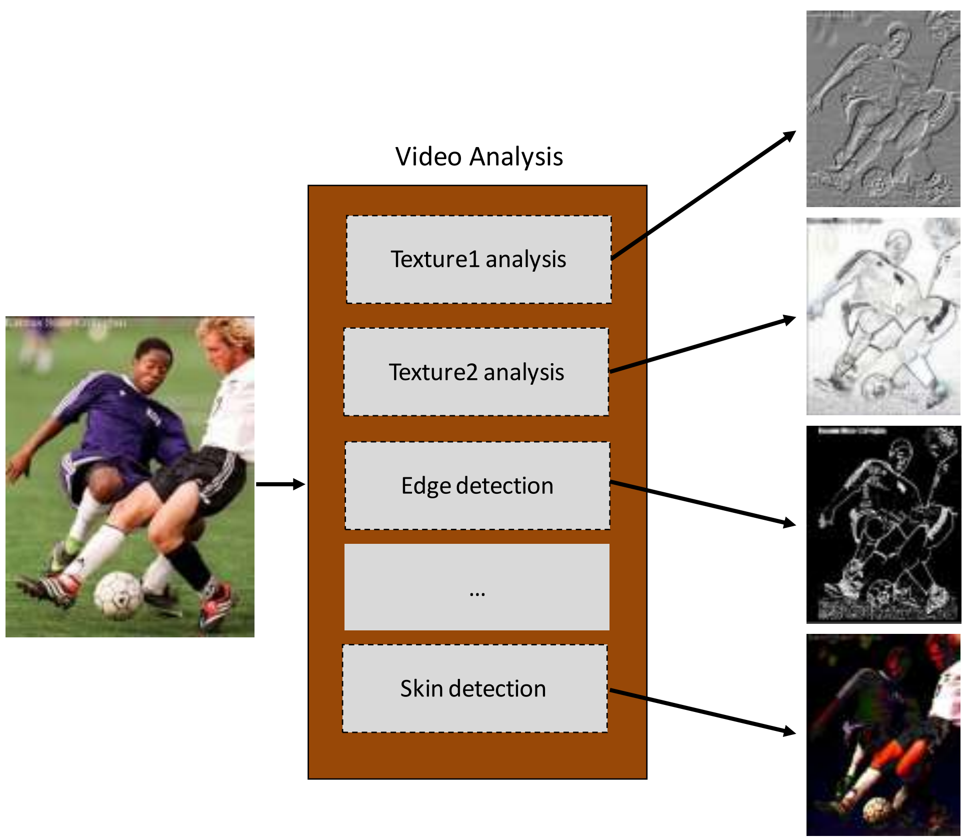

Figure 6 : Video Feature Analysis with some example outputs

H.264 High Profile

H.264 groups compression tools into application related sets which are called profiles. Each profile is optimized for a set of similar applications. For 4:2:0, 8bit video there are 3 commonly used profiles:

H.264 Profile	Description
Baseline Profile	Lowest complexity profile aimed at mobile and other low power devices
Main Profile	Medium complexity profile aimed at standard definition consumer applications
High Profile	High complexity profile aimed at HD and high quality video applications such as HDTV, Blu-Ray, etc

Figure 7 shows the nested architecture of the compression tools for the different profiles. The baseline profile has the smallest (least complex) number of tools and provides the least amount of compression. Main Profile includes all baseline profile tools, and adds others which increase the complexity in return for higher compression efficiency. High Profile continues the trend and provides more tools, higher complexity, and yet more compression efficiency.

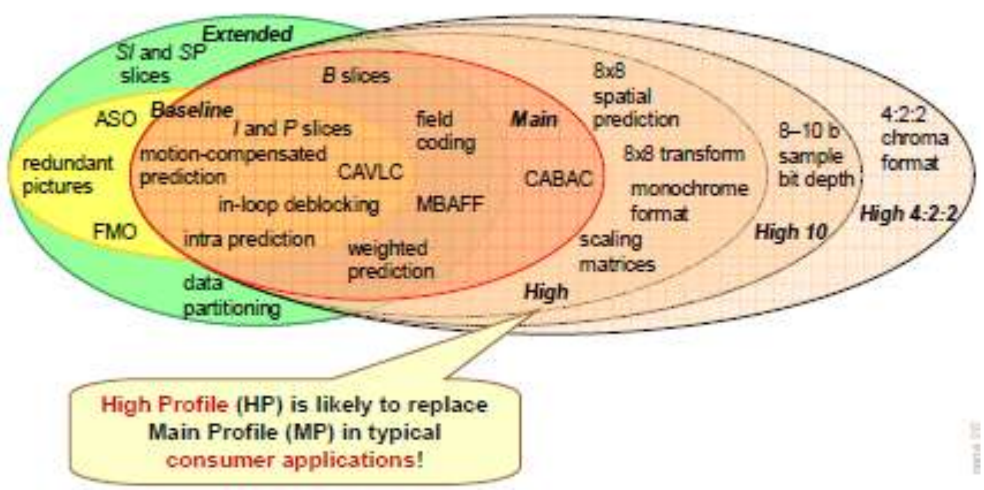

Figure 7 : H.264 Profile Nested Architecture

For SD and HD content, High Profile has become the predominant profile used. High Profile tools provide 15-30% bit rate improvement over Main Profile tools. In addition, HP tools provide subjective video quality improvements that cannot be objectively measured in many cases. This has been demonstrated in various highly regarded viewing tests in the industry including Blu-Ray viewing tests in 2004 and others. The excellent performance of H.264 high profile has resulted in most SD/HD major consumer applications standardizing on it including the Satellite, Cable, Blu-Ray, and other video distribution industries.

Figure 8 shows H.264 video performance for main profile (MP) and high profile(HP). In this example for 1280x720 film content, HP provided 30% improved performance compared to main profile.

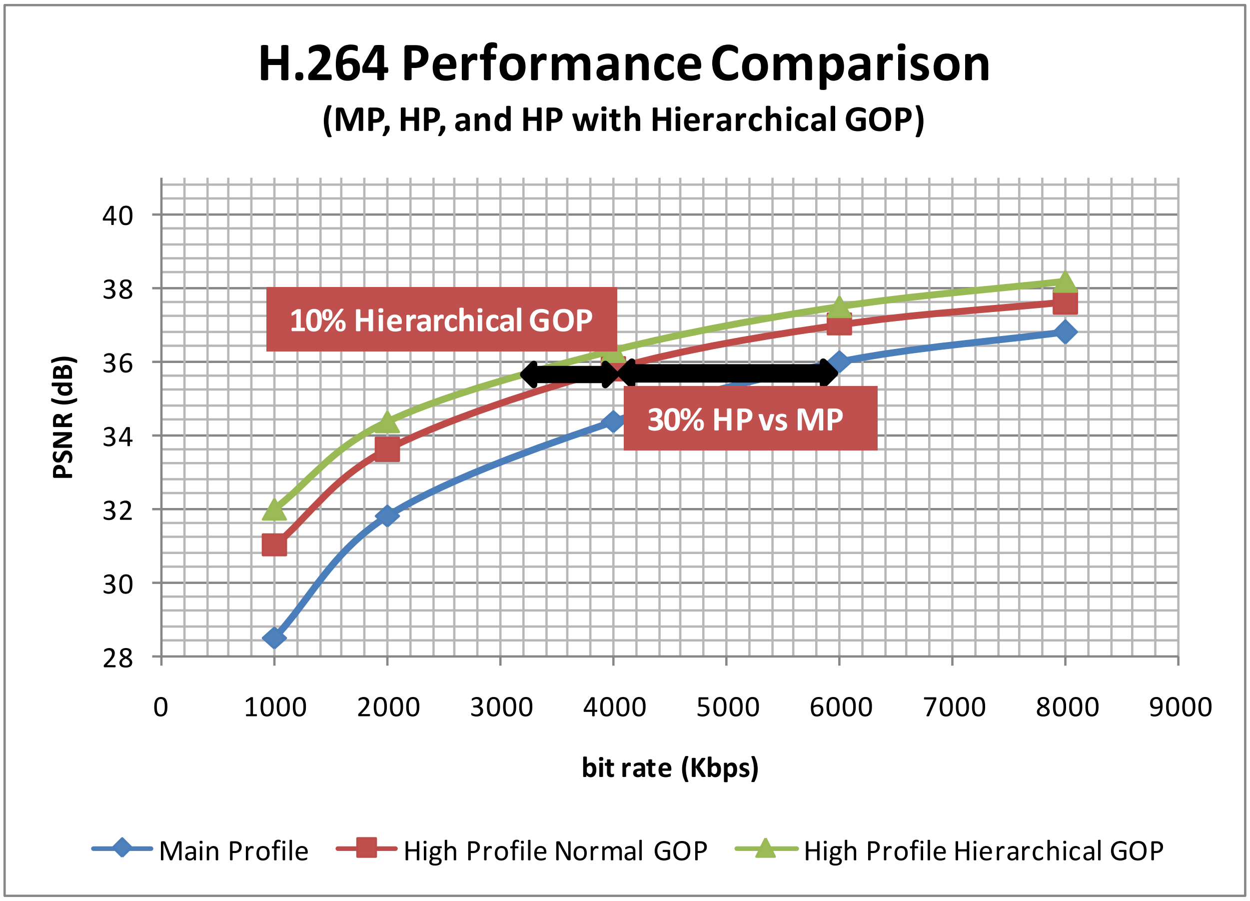

Figure 8 : H.264 MP, HP, and Hierarchical GOP Performance

H.264 Hierarchical GOP Structure

The H.264 video coding standard supports a coding efficiency and scalability tool called hierarchical GOPs. An example GOP structure is shown in the Figure 9. This GOP structure provides coding efficiency by introducing a new frame type named B-reference frames. In normal GOP structures, B frames cannot be used as reference frames, but in a hierarchical GOP structure these new B-reference frames can be used for motion prediction.

Figure 11 shows a typical hiearchical GOP structure. The red and blue frames are B-refererence frames which are used as predicitive frames for the green normal B frames. In this example, the distance between P frames is increased from 3 frames (for a typical IBBP GOP) to 8 frames, reducing the overhead bits from P frame predicition across the GOP.

The result is that GOP structures using more B frames can be constructed, enabling coding gains that B frames have over traditional P frames. A 5-10% compression efficiency gain is typical when using the hierarchical B frames compared to normal GOP structures.

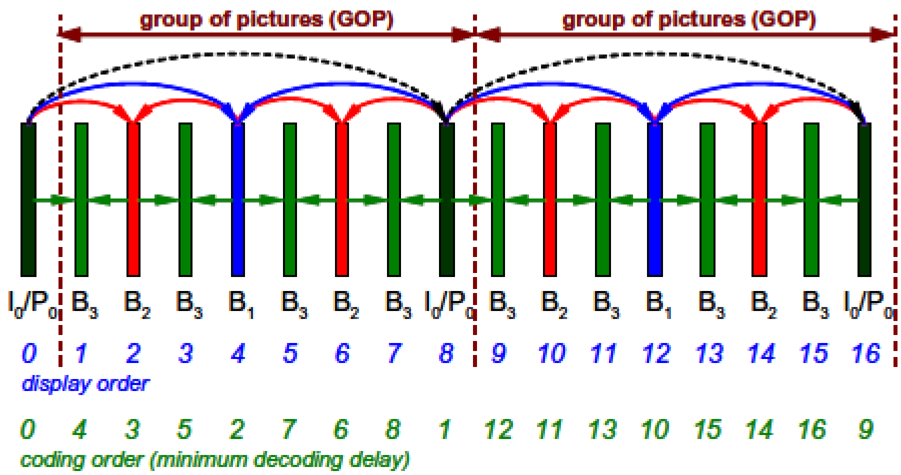

Figure 9 : Hierarchical GOP Example Structure

Another advantage of hierarchical GOPs is that they form a temporal scalable structure. This is the same mechanism as used in the MPEG-SVC standard.

Figure 8 shows H.264 video performance comparison of high profile(HP) and HP with hiearchical GOPs. In this example for 1280x720 film content, HP hierarchical GOP provided 10% improved performance compared to HP with normal GOP structure.

Weighted Prediction

H.264 introduced a compression tool to combat annoying compression artifacts such as blocking, during fade-in, fade-out, and dissolve video sequences. Compression of fading sequences typically results in annoying compression artifacts due to the poor prediction reference signal caused by the fading of the picture. H.264 added the ability to match the fading rate with a prediction signal in the bitstream. In essence, an encoder which can detect a fade or dissolve can apply a uniform fading signal into the prediction signal, greatly improving the video compression performance.

Video compression performance using weighted prediction (WP) provides 20-60% gain during the fade portion of a video signal, and 10-20% gain during a dissolve segment. Of course, these gains are not present during non-fading portions of the video signal. However, due to the very large gain during the fade period, and given the relatively high rate of fades used in commercial video distribution for advertising and the use of fades/dissolves during normal video content creation, the performance benefit is significant.

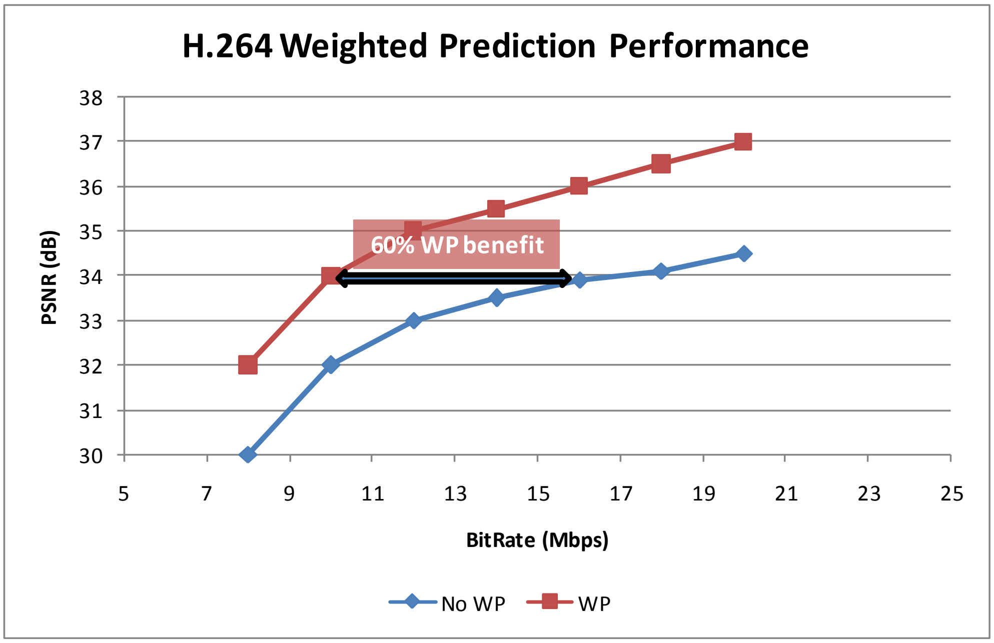

Figure 10 : Weighted Prediction Performance Gain for HD Fade Sequence

Summary of igolgi Video Quality Gains

igolgi's video quality gains compared to a single pass H.264 Main Profile(MP) Encoder are summarized in Figure 11. A range of performance is shown, as video encoder performance can vary with different content. igolgi combined performance gains are typically 25% better video quality. Subjective improvements are higher for some content. Weighted Prediction will add increased performance improvement of typically 10-20% during a fade/dissolve video segment.

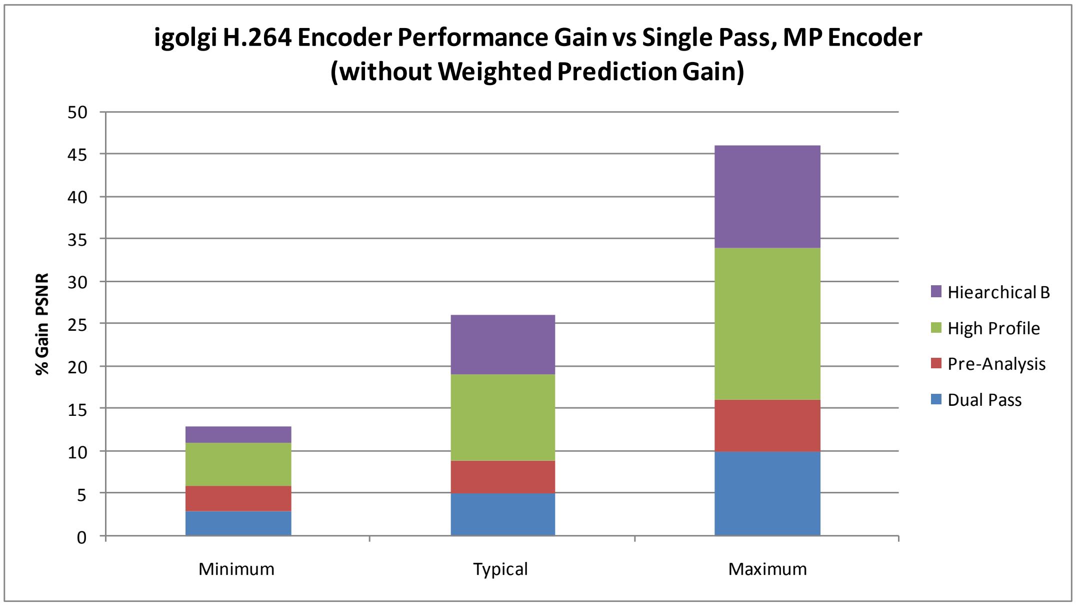

Figure 11 : Summary igolgi Video Quality Gains

Conclusion

igolgi has developed high quality and multi-core scalable software for H.264 and MPEG2 video compression functionality. This results in a fast, real time H.264 High Quality encoder that achieves typically 25% performance gain as compared to a single pass H.264 MP encoder. In addition, igolgi's software solutions have been designed to leverage the CPU trend toward denser multi-core architectures. As these CPUs with increased core density enter the market, igolgi's solutions scale to offer continually lower TCO.