

H.264 Based Video Conferencing Solution

Overview and TMS320DM642 Digital Media Platform Implementation

White Paper

UB Video Inc.

Suite 400, 1788 west 5th Avenue

Vancouver, British Columbia, Canada V6J 1P2

Tel: 604-737-2426; Fax: 604-737-1514

www.ubvideo.com

Introduction

Video conferencing systems are being increasingly used around the world as tools that enable cost-effective and efficient communication while reducing expenses and enhancing productivity. Deployed nowadays mostly within the enterprise setting, video conferencing offers collaboration tools, in addition to the exchange of audio and video information, that make it an attractive alternative to business travel.

Most of the existing systems that are currently deployed run on ISDN infrastructure. However, more and more systems are now enabled to run on either ISDN or IP infrastructure, and it is expected that the majority of the systems that will be deployed over the next few years will run on IP networks. With the continuous adoption of Voice over IP infrastructure within and outside the enterprise, the additional investment needed to provide video over IP is relatively small, implying that the interactive video communication market over IP networks is expected to show significant growth over the next few years.

Most of the existing video conferencing systems operate in the low bit rate range due to the limited bandwidth constraint imposed by the infrastructure used by the corresponding applications. From a video coding point of view, the bandwidth constraint represents an obstacle to the widespread market adoption of video conferencing solutions, since the resulting low bit rate yields an unsatisfactory video quality.

Most of the existing video conferencing systems currently deployed around the world make use of video coding products that are based on either the H.261 and/or the H.263 video coding standards. The emerging H.264 standard is expected to revolutionize the video conferencing industry simply by allowing as much as 50% in bit rate savings over H.263 while maintaining the same or better visual quality. However, H.264 based products have two major issues to deal with. First, encoders and decoders that are compliant with H.264 are expected to be much more complex than the H.261/H.263 counterparts, requiring maximum implementation efficiency. Second, the H.264 standard is still expected to change over the next few months, hence the need for a high level of programmability.

To address the above problems, UB Video has developed UBLive-264BP, an H.264-based video processing solution on the Texas Instruments TMS320C64x platform. UBLive-264BP involves encoding/decoding that is compliant with the H.264 Baseline Profile, as well as pre- and post-processing tools that are intended mainly to improve the visual quality. UBLive-264BP not only achieves real-time performance on the TMS320C64x platform for high-resolution video while maintaining good video quality, but also includes solutions for specific video conferencing problems such as latency and video packet loss. The high-quality and efficient UBLive-264BP, and the high-performance TMS320C64x family of DSPs, represent a compelling integrated software/hardware solution for the next generation video conferencing applications.

The purpose of this paper is to present, from a video coding perspective, the challenges that are currently facing video conferencing applications, as well as a discussion of UBLive-264BP, the UB Video solution and its implementation on the TMS320C64x family of DSPs. First, an overview of the video conferencing market is given, followed by a more detailed discussion of the technological challenges of video conferencing applications. UBLive-264B is then discussed and its implementation on the TMS320C64x family of DSPs is presented.

Video conferencing Market: Overview

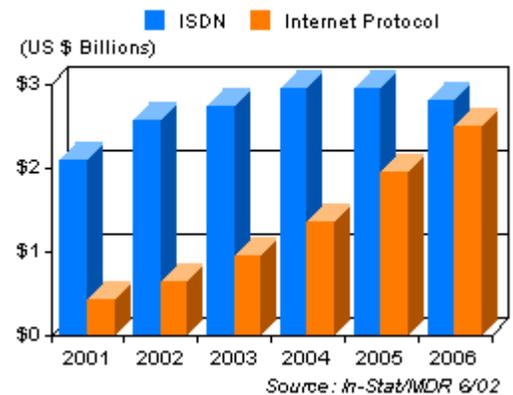
Ever since AT&T unveiled the first videophone nearly four decades ago, much effort has been made towards achieving the ultimate objective of establishing visual communications within the enterprise and progressively making such a dream a reality within the consumer market. Today, video conferencing has made many strides in that direction and companies are taking a hard look at the technology not only as a cheaper alternative to business travel, but also as an important tool for the day-to-day operations.

The majority of video conferencing sessions take place in an enterprise setting over an ISDN or IP-based network. The ISDN technology has been around for more than a decade, though its growth has recently stalled as IP technologies have gained both momentum and wider acceptance. At this time, more than 80% of the video conferencing systems support both the ISDN and IP modes of operation. Users of video conferencing equipment are in the process of converting their existing ISDN infrastructure over to IP, and the trend indicates that in the next 5 to 10 years, most video conferencing systems will be IP based. While IP is relatively inexpensive and Ethernet connections can be found pretty much everywhere in the corporate setting, IP also does not provide quality-of-service guarantees.

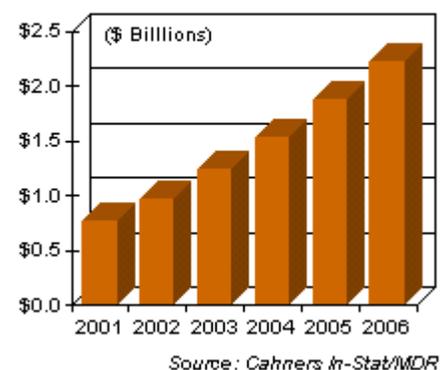
The video conferencing market is divided into three categories: endpoints, infrastructure and service providers. The endpoints market is segmented into three sectors: enterprise group systems, enterprise personal systems and consumer systems. The infrastructure market consists of Gateways, MCUs as well as Gatekeepers. The gateways (a.k.a network bridges) allow connections between different ISDN and IP network types through trans-coding and/or trans-rating. An MCU can combine multiple incoming calls into a single conference. Some higher-end group conferencing systems come with built-in MCUs that support up to four additional users. Service providers and organizations with large numbers of IP-based endpoints use gatekeepers. Gatekeepers can be used to reserve and set up conference calls, as well as limit the number of simultaneous calls to maintain high-quality individual calls. Video conferencing service providers deliver real-time video, voice and data, and are expected to drive the convergence of video conferencing services, instant messaging and web collaboration, among other services.

Existing video conferencing products are designed mostly for the enterprise market. Analysts anticipate that such products will penetrate the consumer market at the end of this decade. This will likely not happen without the adoption of the video conferencing technology and its integration into existing and new solutions by large telecommunication companies such as AT&T, MCI and Sprint. Evidence exists that such companies are inclined to do just that fairly soon (if not already). However, the other bigger uncertainty is whether the consumer will ever adopt an appliance such as the videophone or use the PC to conduct the video conferencing session. A recent In-Stat/MDR survey indicates that attitudes towards the use of, and participation in, video conferencing

Worldwide Total Video Conferencing Services Revenue Forecast - ISDN versus IP



Forecast for Worldwide Video Conferencing Systems Market



activities are changing in a positive way. For example, the total projected market for video conferencing systems is expected to increase from \$1 Billion in 2002 to \$2.2 Billion by 2006. In addition, desktop PC video conferencing systems are expected to re-emerge in the next few years. North America represents, by far, the largest market for video conferencing equipment, and with an expected significant growth during the coming 2 years. Other countries, particularly in Europe and Asia, have shown significant market growth opportunities as well.

From the video coding perspective, there is currently a shift taking place from hardware-based codecs to software-based codecs. The latter are expected to dominate the endpoint market. . The main reason for the shift is that software-based codecs can be easily customized to accommodate different needs, such as for example the need to support different video and audio coding formats as well as the need to customer-specific functionality.

Challenges of video conferencing applications: A video perspective

The widespread adoption of video conferencing solutions would not become a reality without the development of effective solutions to many problems in areas such as the transport of audio and video data and the management of the communication infrastructure. From a video coding and transport perspective, a number of issues need to be addressed in order to ensure the efficient utilization of the available bandwidth and to guarantee a good video quality. These challenges are discussed next.

◆ **Efficient bandwidth utilization**

Video conferencing applications usually involve the exchange of audio, video, and data between participants in the conferencing session. However, most of the bandwidth allocated to a typical video conferencing session is usually consumed by the video data. For example, in a typical video conferencing session running at 384 kbps, the audio and other non-video data may need a bit rate of 80-100 Kbps, leaving at most 300 Kbps for video. Any significant reduction in the bandwidth required to maintain a given desired video quality could result in a number of benefits, such as:

- Increasing the video quality for a given rate.
- Increasing the number of participants during a conferencing session and that could be accommodated simultaneously during a conference call, and
- Increasing the amount of data, other than audio and video data, which could be exchanged during the conferencing session.

Moreover, most of the video conferencing applications are usually allocated a relatively low-bandwidth. This is possible mainly because typical video conferencing content consists mostly of a static background and slowly varying foreground (typically head and shoulders). A typical conferencing session where the video resolution is 352x288 may run at 384 kbps or even lower bit rates such as 256 Kbps, with a relatively acceptable video quality. The fact that video conferencing applications run at low rates necessitates the use of the most effective video compression tools to maintain a good video quality.

♦ Low processing delay

Real-time video conferencing applications involve the processing and transmission of mostly timely information, represented by the audio and video data. Given that participants are in a constant exchange of audio and video information, it is very important that the delay associated with the processing and transmission of the data be kept to a minimum, while also maintaining good quality. The delay, also referred to as latency, can be divided into three main components: encoding delay, network delay, and decoding delay. In real-time interactive applications, the user will notice some objectionable delay if the round trip time exceeds 250ms. The processing delay at both the encoder and decoder represents a significant component of the overall delay, with the encoder delay being at least twice as much as the decoder delay. In order to maintain a minimal latency, it is therefore important to maintain a very small processing delay.

♦ Better Video Quality

The video quality in conferencing applications can be affected negatively by a number of factors, including noise and brightness changes in the source video, the presence of trailing artifacts in the reconstructed video, and network packet loss. Each of these issues is discussed in detail in the following sections.

Noise and Brightness Changes

Given the low bit rate constraints in video conferencing applications, it is desired to minimize the amount of information being transmitted over the communication channel. The original video frames captured by a camera during a conferencing session are usually affected by temporal and spatial noise. The noise is much more perceived when the lighting conditions are not adequate. Moreover, depending on the conferencing setting and the lighting conditions, brightness changes may frequently be introduced in successive video frames during the conferencing session. For example, in the case of a video conferencing session using a webcam, when a person is very close to the camera, significant brightness changes occur often particularly if the person moves frequently across the field of view of the camera.

The presence of noise in the video frames and the brightness changes between successive video frames can degrade the efficiency of inter prediction. This is due to the fact that motion compensated prediction, the most widely used approach for inter prediction, assumes that the moving object would have the same luminance values in two successive frames. When the luminance values change due to noise or brightness changes, the motion estimation algorithm in the encoder is likely to interpret the changes as being due to motion. This would increase substantially both the computation and bandwidth requirements, and would also lead to degradation in video quality (perceived mostly in terms of annoying blocking artifacts). Consequently, it is necessary to apply pre-processing methods to minimize the effects of both the noise and the brightness changes on video quality.

Coding Artifacts

Since most of the video conferencing sessions are usually held at low bit rates, the resulting video quality may be low. One common consequence of the low bit rate constraint is the presence of trailing artifacts that appear as a series of dots trailing the moving object. The appearance of trailing artifacts is due to aggressive coding at such low bit rates.

◆ **Error Resilience**

The compressed bit stream generated by the video encoder is first segmented into fixed or variable length packets. It is then multiplexed with audio and other data, and the resulting packets are then sent to the network. It often happens that some of the transmitted packets get lost or corrupted during transport over the network, due for example, to congestion in the network. In some video conferencing sessions taking place over wireless links or over the Internet, up to 30% of the packets can get corrupted, leading to a large degradation in audio-visual reproduction quality. Two methods can be developed to address the above-mentioned problems. The first method, also known as the error resilience method, is to introduce redundancy in the transmitted data and/or to reduce the dependencies among the adjacent transmitted pieces of data. The second method, also known as the error concealment method, aims at hiding as much as possible the effects of the detected errors in the video data in the reconstructed video frame. Since a typical video conferencing system may include equipment from different vendors, it is often assumed that the decoder does not perform error concealment, and it is therefore necessary to employ appropriate error resilience tools at the encoder.

UBLive-264B Solution

UB Video's UBLive-264BP is a complete video processing solution that has at its core an optimized H.264 encoder/decoder. The UBLive-264BP encoder and decoder are based on the Baseline Profile of H.264, which has a number of features that make it suitable for video conferencing applications. The main features of the H.264 Baseline encoder/decoder are described in detail in [1]. UBLive-264BP was designed to address all of the major video processing issues that plague video conferencing applications, as explained in the following.

◆ **Up to 50% in bit rate savings**

Most of the existing video conferencing systems in existence today are based on video coding solutions that make use of either the H.261 or H.263 video coding standards. Compared to H.263-based solutions, UBLive-264BP yields an average reduction in bit rate by up to 50% for a similar degree of encoder optimization at most bit rates and for the same subjective video quality. The savings in bit rate are due in part to the flexibility provided by H.264.

◆ **Low Processing Delay**

UBLive-264BP also features very low processing latency, a key requirement in video conferencing applications. UBLive-264BP guarantees an excellent video quality even for single-pass encoding, thereby reducing processing latency. The reduction in latency is due to the collection of complexity-reduction techniques UB Video has researched and developed over the past few years.

◆ **Better Video Quality**

UBLive-264BP yields high video quality through the effective use of the H.264 Baseline Profile features as well as the use of pre-processing methods such as motion-compensated filtering, brightness compensation in the source video, and trailing artifact avoidance/reduction, as well error resilience methods such as Flexible Macroblock Ordering (FMO) and the use of redundant slices.

Motion Compensated Noise Filtering

In motion-compensated filtering, a source pixel is filtered by taking into account its trajectory in previous video frames. Motion compensated filtering is computationally demanding and can prove very computationally demanding in real-time video conferencing applications. However, UB Video's UBLive-264BP includes a very efficient motion compensated filtering method that provides a good balance between savings in bit rate, number of processing cycles and video image quality.

Global Brightness Change Compensation

Global brightness changes that take place during video conferencing sessions are mostly due to camera adjustments, and they most often do not need to be transmitted and reproduced at the decoder side. The UBLive-264BP solution includes a brightness change compensation filter that maintains the brightness level constant during the conferencing session. The brightness compensation filter estimates the amount of change in brightness between the previously coded frame and the current source frame, and compensates the source frame to bring its brightness level close to that of the previously coded frame. Consequently, no additional bits are needed to transmit the brightness change, and the overall video quality is improved.

Trailing Artifact Removal

The presence of the trailing artifacts in video frames can significantly degrade the quality of the resulting video images. UBLive-264BP includes in the encoding algorithm a trailing artifact avoidance filter that checks for the existence of those artifacts in the identified areas and provides solutions to avoid the detected artifacts.

Flexible macroblock ordering (FMO)

Macroblocks in a given frame are usually coded in a raster scan order. With FMO in UBLive-264BP, on the other hand, macroblocks are coded according to a macroblock allocation map that groups, within a given slice, macroblocks from spatially different locations in the frame. Such an arrangement enhances error resilience in the coded bit stream since it reduces the interdependency that would otherwise exist in coding data for adjacent macroblocks in a given frame. In the case of packet-loss, the loss is scattered throughout the picture, and errors can therefore be easily concealed.

Redundant slices

This UBLive-264BP feature allows the transmission of duplicate slices over error-prone networks to increase the likelihood of the delivery of a slice that is free of errors.

UBLive-264BP: Demonstration Software

UB Video has demonstration software for UBLive-264BP that runs on the TMS320DM642 Network Video Developer's Kit (NVDK) board from Texas Instruments.

The TMS320DM642™ digital media processor (DM642) DSP core delivers the highest performance at the lowest power consumption of any available DSP in the market to date. With its 600Mhz processing power in production today and an aggressive process technology roadmap for continued higher clock rates, this DSP is most suited to overcome the complexity and computational requirements of H.264 and to deliver high-quality full-screen video for most broadband video applications. The DSP core processor has 64 general-purpose 32-bit registers and eight highly independent functional units - two multipliers and six arithmetic logic units

(ALUs) with VelociTI.2 extensions. The VelociTI.2™ extensions in the eight functional units include new instructions to accelerate the performance in video and imaging applications and extend the parallelism of the VelociTI™ architecture. The DM642 uses a two-level cache-based architecture and has a powerful and diverse set of peripherals. The Level 1 program cache (L1P) is a 128-Kbit direct-mapped cache and the Level 1 data cache (L1D) is a 128-Kbit 2-way set-associative cache. The Level 2 memory/cache (L2) consists of a 2-Mbit-memory space that is shared between program and data space. L2 memory can be configured as mapped memory, cache, or combinations of the two.

The UBLive-264BP demonstrates the ability to do real-time H.264 encoding and decoding with a feature set optimized for low-delay coding. The demonstration runs on the TMS3200C64x Network Video Developer's Kit (NVDK) from Texas Instruments that includes the TMS320C6415 DSP. A block diagram of the demonstration is shown in Figure 1. Video is captured from a camera and (optionally) scaled to SIF resolution. The video is then real-time encoded at the bit rates of 128 and 384 Kbps for SIF content and 1 Mbps for full-screen content, and the bit stream is then fed back to a real-time decoder. The DSP applies format (4:2:0 to 4:2:2) and color conversion to the decoded video before display. The encoder and decoder modules have a documented API that makes it easy to integrate the modules in end equipment products.

The UBLive-264BP software illustrates that UB Video's Baseline H.264 solution allows a 35-50% bandwidth reduction over that required by H.263-based solutions. Using the video quality enhancement tools available in UBLive-264BP, the resulting video quality at such a rate is similar to or even better than that provided by H.263 based solutions. And while the additional complexity is perceived to be high, the software also demonstrates that UBLive-264BP can simultaneously encode and decode 640x480 interlaced video sequences on a single 600Mhz DM642 chip, with possibly enough room for audio and other video conferencing system components.



Figure 1. Block diagram of the UBLive-BP demonstration.

For more information on UBLive-264BP, please contact UB Video at www.ubvideo.com.

References

[1] Emerging H.264 Standard - Overview and TMS320C64x Digital Media Platform Implementation. UB Video Inc. www.ubvideo.com.