Overview of the Grand Alliance HDTV Video Compression System

Joel Zdepski, David Sarnoff Research Center
Regis Saint Girons, Thomson Consumer Electronics
Paul Snopko, Zenith Electronics
Vinay Sathe, General Instrument
Woo Paik, General Instrument
Eric Petajan, AT&T
Jae Lim, MIT
Xavier Lebegue, Thomson Consumer Electronics
Kiran Challapali, Philips Electronics North America

Abstract

The US HDTV process has fostered substantial research and development activity over the last several years. Four digital HDTV systems where tested at the Advanced Television Testing Center for consideration as the US HDTV standard, however a single system was not selected. A consortium of the four digital proponents, called the Grand Alliance (GA), announced a combined system and submitted it to ACATS for consideration. This paper describes the video compression subsystem and the hardware prototype.

1. Introduction The competitive phase of the HDTV standardization process produced some innovative video compression algorithms [1].. The systems where implemented in real-time hardware and received a rigorous evaluation not possible through computer simulation and paper analysis. The core compression processing used in all of the systems was motion compensated temporal prediction followed by spatial decorrelation with the Discrete Cosine Transform (DCT) [2]. In each system, the resulting transform coefficients were quantized and entropy coded with a Huffman code. The main areas of difference were in picture format, modes of temporal prediction, and the transmission system. The GA system has been designed to incorporate most key video compression technology utilized in previously tested systems, and is designed to perform well with both progressive and interlaced sourcee pictures. We have emphasized compatibility with MPEG (Moving Picture Experts Group) syntax [3]. The Grand Alliance system conforms the Main Profile of the MPEG-2 standard, implemented at the High Level [4-5].

2. Compression for US HDTV application

Modern digital transmission technologies deliver approximately 17 Mbps to about 20 Mbps within a single 6 MHz terrestrial channel. This means that encoding the HDTV video source requires a bit rate reduction by a factor of 50 or higher. The methods utilized, for example, include source adaptive processing, motion estimation/ compensation, transform domain data representation, and statistical coding. The GA encoder system is shown in Figure 1. The audio and video source is encoded by the corresponding encoder, the output of which is a string of bits that represents the source. A higher layer of multimedia synchronization coding is added by the PES layer which is further multiplexed by the transport layer. Arbitrary data signals can also be mixed with the compressed HDTV signals. The channel coder transforms the string of bits to a form suitable for transmission over a channel through some form of modulation, here represented by the 8-VSB modem. The modulated signal is then transmitted over a communication channel.

Figure 2 shows the functional block diagram of the GA video encoder. It consists of a preprocessor, a motion estimation subsystem, a DCT and Quantization sub system, and a variable word-length encoder subsystem. Finally, a Rate-Buffer is required to smooth the data for constant bitrate transmission through the GA Transport and Transmission system. In this section we describe each component of the video compression encoder.

2.1 Video Preprocessor A key feature of the GA system that helps meet this challenge is the

1058-6393/95 $4.00 © 1995 IEEE

adaptive use of multiple formats. Namely, the GA system will display video at frame rates related to frequencies of 59.94 and 60.0 Hz, and two families of picture formats based on 1280 (H) x 720 (V) and 1920 (H) x 1080 (V) active pixels have been chosen. One consequence of incorporating multiple format capability is the need for either multisync displays or electronics in a transconversion subsystem in the decoder to drive the decoders native display format.

The input video source to the GA video compression system is in the form of R (red), G (green), and B (blue) components. To reduce the correlation between the RGB signals and to exploit the response the human visual system, the RGB components are converted to the YC_1C_2 color space through a linear transformation. Y corresponds to the luminance (intensity or black-and-white picture) while C_1 and C_2 correspond to the chrominance. Furthermore, the human visual system is less sensitive to high frequency components of the chrominance components than of the luminance component. Consequently, the chrominance components are low-pass filtered and sub-sampled by a factor of two along both the horizontal and vertical dimensions.

2.2 Motion Estimation and Compensation The GA compression algorithm utilizes multiple prediction methods to effectively provide motion compensation. The prediction methods include frame prediction, adaptive field prediction, dual prime (used for forward prediction only), and bi-directional prediction.

Motion estimation processing is the primary contributor to encoder complexity. The main factor is the size of the region in the prediction frame to be searched. The MPEG-2 High Level limits the search displacements of -2048 to +2047.5 pels horizontally, and -128 to +127.5 pels vertically. Since an exhaustive search of these area at the 1/2 pel level would require approximately 2^{28} pixel difference calculations/MB. Clearly more efficient algorithms or restricted search ranges are required in today's practical encoders.

2.3 Refreshing Options A motion compensated prediction loop is not practical without some form of refreshing. Decoder initialization occurs after the channel is changed, or the signal is lost and then recovered. The GA system satisfies the need for refreshing with two different mechanisms. I-frame refreshing uses periodic intra-coded frames. Progressive refreshing uses periodic intra-coded (16x16) macroblocks. In either case, each MB position must be refreshed at least once within a 132 coded frames.

2.4 Adaptive Field/Frame Processing For interlaced video signals the encoder may either separate each frame into its two field components and then process the two fields independently, or it can process the two fields as a single frame by interleaving the lines of field 1 and field 2. Field processing generally works better than frame processing in detailed moving areas, while Frame processing works better than field processing when there is little or no motion. The GA system combines the advantages of both frame processing and field processing by adaptively selecting one of the two modes on a block-by-block basis.

2.5 Forward Estimation for Rate Prediction Motion compensation, adaptive quantization and variable length coding produce highly variable amounts of compressed video data as a function of time. Therefore, compressed video data buffering in the encoder and decoder is required for efficient channel utilization. Buffer size is constrained by the maximum tolerable delay through the system for a particular application and cost. The fullness of the buffer is controlled by adjusting the amount of distortion or quantization error in each image. A feedback control system is required to regulate the distortion level which controls the buffer fullness to prevent overflow. The design of this control system is complicated by:

- Delay between a change in the distortion level and the subsequent change in buffer fullness.
- Perceptual constraints on the instantaneous and average distortion levels.
- Difficulties in modeling the bit-rate as a function of distortion level.

The visibility of distortion due to an increase in scene complexity is minimized by smoothly increasing the distortion level. In the GA rate control system, the complexity estimate is the variance of the ideal (unquantized) displaced frame difference (DFD). Since the whole-pel motion vectors are available one frame ahead of the loop, the ideal displaced frame difference is calculated using whole-pel accuracy (i.e., no sub-pel estimation). The pixel values of the ideal DFD are squared and summed over the picture.

2.6 Adaptive Intra and Non-Intra Quantization Matrices The MPEG-2 syntax allows the quantization matrices to be specified for every picture for improved coding efficiency. A certain probability distribution is associated with the VLC codes for quantized coefficients in the MPEG standard. Although one cannot change the VLC distribution to match the actual distribution of the data, the quantization matrices can be adjusted to help match the distribution of the data to the distribution of the VLC.

2.7 Perceptual Coefficient Selection The human visual system is not uniformly sensitive to coefficient quantization error. The amount of visible distortion resulting from quantization error for a given coefficient depends on the coefficient number or frequency, the local brightness in the original image and the duration or the temporal characteristic of the error. Low frequency distortion is more visible than higher frequency distortion which appears as noise or texture.

Displays and human visual systems exhibit non-uniform sensitivity to detail as a function of local average brightness. Loss of detail in dark areas of the picture is not as visible as it is in brighter areas. Another opportunity for bit savings is presented in textured areas of the picture where high frequency coefficient error is much less visible than in relatively flat areas. Brightness and texture weighting require analysis of the original image since these areas may be well-predicted in the DFD. Finally, distortion is easily masked by limiting its duration to one or two frames.

Perceptual Selection is used in the GA system using the properties of the human visual system to code pictures using fewer bits within a perceptually consistent level of quality. For each transform block, the perceptual selection method determines the acceptable amount of distortion per each frequency band; if the magnitude of the coefficient in the loop is smaller than the acceptable distortion level, then the coefficient is set to zero, regardless of the quantization step size level.

2.8 Adapting M and N The motion estimator range of practical encoders is often reduced for complexity considerations. The use of B-frames reduces the effective motion tracking range by a factor of M where M-1 is the number of B-frames between a given pair of I- or P-frames. Scenes with rapid large area motion may exceed the range of the motion estimator if M is too high. This condition is detected in the forward analyzer which then signals the frame reorder section and motion estimator to use a lower M value for subsequent frames. The increased bit-rate caused by the temporary motion estimator overrun is absorbed by the rate buffer to maintain picture quality.

2.9 Film Mode The GA system is able to detect source material originating from 24 fps film that has been removes the redundancy and removes them to returnto the original 24 fps source format prior to video encoding. The removal of redundancy ensures that the same information is not transmitted repeatedly, thus significantly improving the efficiency of the video compression system.

3. Transport and Modem Subsystems Figure 1 illustrates the position of the video encoder with respect to the transport and modem subsystem in the overall system. The Transport consists of the PES encoder and the Transport Multiplexer. The Grand Alliance specification has made use of portions of the MPEG-2 Systems specification [6]. The transport resides between the application (e.g. audio or video) encoding/decoding function and the transmission subsystems. The transport system consists of two logical layers, the Transport Layer and the packetized elementary stream (PES) layer. The transport layer is a multiplex of fixed length packets, each 188 bytes long. The single transport packet consists of a 4 byte header and a 184 byte payload area. It provides both Transport and Link level functionality in the Grand Alliance system. The Header is used to demultiplex the elementary streams. The Payload region of the packet carries the elementary stream data. The PES packet header carries information for the application which allows the presentation of audio and video to be synchronized.

Following Transport encoding, the digital data are processed by the Grand alliance transmission subsystem. The digital data, generated by video and audio encoders are D/A converted into 8-level symbols (3 bits/symbol) which are vestigial sideband (VSB), suppressed carrier, AM modulated and transmitted in a 6 MHz channel [7]. The system is referred to as "8-VSB". The 8-VSB transmission systems incorporates several improvements over previously tested systems, increasing the data rate while maintaining robust terrestrial transmission performance. The symbols are arranged in "Data Frames" each comprising 313 "Data Segments" of 832 symbols Each Segment starts with a constant 4 symbol segment sync (SS) and each Frame starts with one full segment of Field Sync (FS) consisting of a sequence of pseudo-random symbols. Each payload segment ends with 80 symbols of Reed-Solomon forward error correction (RS-FEC).

4. Grand Alliance prototype video compression hardware A prototype video coder-decoder (codec) that demonstrates the salient features of the GA video compression specification (as described in this document) will be built. This section describes some of the features. The coding parameters and algorithms of the compression system are being fine tuned through simulation of the system in software.

The video codec will be able to compress multiple video formats depending upon the source (both interlaced and progressive). The compression technique is essentially independent of the video format, although the coding parameters change. The codec will operate at compressed bit- rates of around 17 Mbps.

Each video frame is coded in one of three modes: Intra-coding (MPEG I-pictures), Predictive coding (MPEG P-pictures) or Bi-directional predictive coding (MPEG B-pictures). The period of I-pictures (Group-of-pictures size, N) and the distance in number of frames between two anchor (I or P) pictures (known as M), can be programmed.

Motion of up to one-half pixel accuracy will be estimated. Integer pixel motion will be estimated on the original video frames, while the sub-pixel refinement will be done on reconstructed pictures. A search range of up to ± 32 pixels vertically and ± 128 pixels horizontally will be used for P-pictures and ± 32 pixels vertically and ± 64 pixels horizontally in B-pictures.

The encoder detects "three-two pulled-down" (originally from a 24 fps movie source) and removes the redundant fields/frames in the sequence. The decoder decodes the pictures and displays them as three-two pulled down material.

Interlaced pictures can be predicted and coded adaptively as fields or frames on a local (macroblock) basis. The adaptive field/frame motion prediction decision will be based on mean absolute error criterion, whereas, the adaptive field/frame DCT will be based on quantized errors or mean absolute error criteria.

Asophisticated adaptive quantization and rate control algorithm will be used. This algorithm will be implemented on a DSP-based subsystem. In addition, a forward analyzer (to assist the rate controller) will be implemented. Coding of the quantized coefficients is in accordance with the MPEG-2 standard.

In summary, the GA prototype hardware video codec will compress and code video in multiple raster formats. The compressed bitstream produced will conform to the GA video coding specification [4]. The bitstream produced will also be compatible with the MPEG-2 video coding standard.

5. Conclusions. The Grand Alliance system provides an excellent solution for terrestrial transmission of HDTV. The constituent companies obtained valuable experience during the first phase of FCC testing, achieving excellent picture quality over terrestrial broadcast channels. The Grand Alliance system incorporates the best features of the individual systems.

The MPEG-2 compression syntax is used as the basis of the video coding. The GA system uses the Main Profile tools and a picture format which conforms to the High Level. With a data rate of approximately 18 Mbps for video coding, compression ratios of approximately 0.35 bits/pixel must be achieved. Source adaptive processing is used to optimize performance of compression while allowing multiple video formats.

6. References

1. FCC Advisory Committee on Advanced Television Service, *Federal Communications Commission Advanced Television System Recommendation.* IEEE Transactions on Broadcasting, March 1993. **Vol 39**(No 1): p. pp. 3-245.
2. Hopkins, R., *Progress on HDTV broadcasting standards in the United States.* Signal Processing: Image Communications, 1993. **Vol 5**: p. pp. 355-378.
3. ISO/IEC, MPEG-2 Video DIS 13818-2, 1994.
4. Grand Alliance, *Grand Alliance HDTV System Specification*1994, Submitted to FCC Advisory Committe on Advanced Television Service:
5. Challapali, et.al., "The Grand Alliance System for US HDTV", To be published in Proceedings of the IEEE, Feb. 1995
6. ISO/IEC, MPEG-2 Systems DIS 13818-1, 1994.
7. Citta, R., *et al.*, *The Digital Spectrum Compatible HDTV Transmission System.* IEEE Transactions on Consumer Electronics, August 1991. **Vol. 37**(No. 3): p. pp. 469-475.

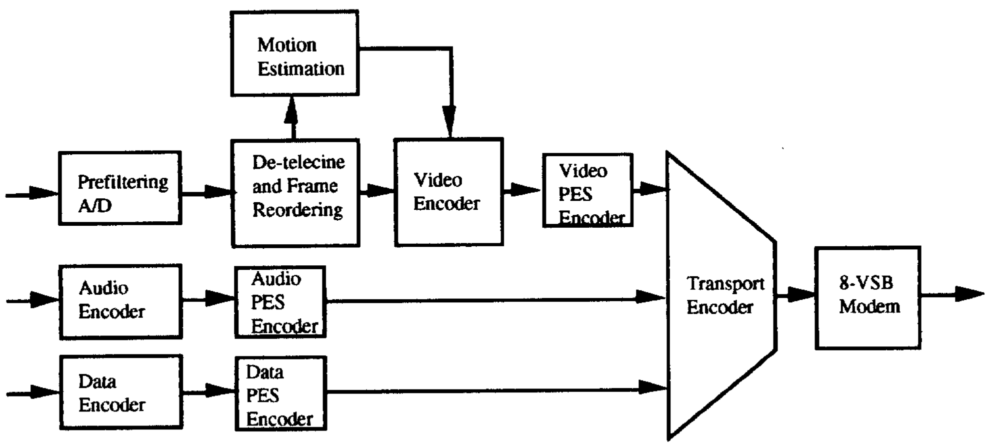

Figure 1 Grand Alliance Encoder System Block Diagram.

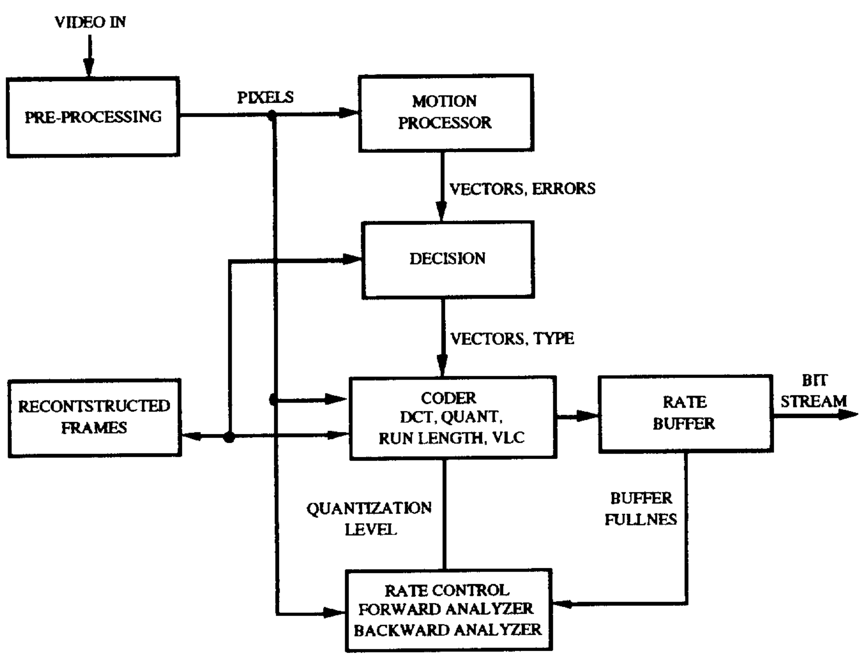

Figure 2. GA Video Encoder