Chapter 2. Advanced Telecommunications and Signal Processing Program

Academic and Research Staff

Professor Jae S. Lim

Visiting Scientists and Research Affiliate

M. Carlos Kennedy

Graduate Students

David M. Baylon, Rubén E. Galarza, Theresa Huang, Raynard O. Hinds, Eric C. Reed, Wade K. Wan

Technical and Support Staff

Cindy LeBlanc, Christopher R. O'Donnell

2.1 Introduction

The present television system was designed nearly 50 years ago. Since then, there have been significant developments in technology, which are highly relevant to the television industries. For example, advances in the very-large-scale-integration (VLSI) technology and signal processing theories make it feasible to incorporate frame-store memory and sophisticated signal processing capabilities in a television receiver at a reasonable cost.

To exploit this new technology in developing future television systems, the Advanced Television Research Program (ATRP) was established at MIT in 1983 by a consortium of U.S. Companies.

The major objectives of the ATRP are:

- To develop the theoretical and empirical basis for the improvement of existing television systems, as well as the design of future television systems.
- To educate students through television-related research and development and to motivate them to undertake careers in television-related industries.
- To facilitate continuing education of scientists and engineers already working in the industry.
- To establish a resource center to which problems and proposals can be brought for discussion and detailed study.
- To transfer the technology developed from this program to the industries.

The research areas of the program focused on a number of issues related to digital television design. As a result of this effort, significant advances have already been made, and these advances have been included in the U.S. digital television standard. Specifically, the RLE Advanced Telecommunications and Signal Processing group represented MIT in MIT's participation in the Grand Alliance which consisted of MIT, AT&T, Zenith Electronics Corporation, General Instrument Corporation, David Sarnoff Research Center, Philips Laboratories, and Thomson Consumer Electronics. The Grand Alliance digital television system served as the basis for the U.S. Digital Television (DTV) standard, which was formally adopted by the U.S. Federal Communications Commission in December 1996.

In addition to research on issues related to the design of digital television system, the research program also includes research on signal processing for telecommunications applications and research on speech enhancement.

2.1.1 Patent

Lim, J.S. *Advanced Television System Using A Different Encoding Technique for Non-Image Areas*. U.S. Patent No. 5,771,073

2.2 Study on Migration to Higher Resolution Digital Television Systems

Sponsor

Advanced Telecommunications Research Program

Project Staff

Wade K. Wan

High-definition television (HDTV) has already begun broadcasting throughout the United States. The formats in the current standard have limited resolution and the need to broadcast at higher resolutions has already been recognized. A generally accepted goal for future terrestrial broadcasting is more than 1000 lines with progressive scanning at 60 frames per second within a single 6 MHz channel. Current transmission and compression technologies cannot support such a high video resolution. We are currently investigating methods to transmit video at the desired resolution goal using enhancement data. This approach would consist of sending two sets of information within the bandwidth allocated for HDTV transmission. Standard video bits would be transmitted at a resolution format allowed by the current digital television standard. The standard video bits will only use part of the bandwidth allocated for transmission and the remainder will be allocated to enhancement data bits. An advanced receiver could receive the standard video bits and then convert the video to a higher resolution format with the assistance of the video enhancement bits. Standard receivers could ignore the enhancement bits. This backward-compatible approach is highly desirable so as not to render earlier receivers obsolete.

An issue that is currently being examined is the selection of the resolution format that will be used to generate standard video bits. This selection will affect the video enhancement bits. For example, 720-line, progressively scanned, 60 frames per second video may be transmitted and then spatially upsampled at the receiver. Alternatively, 1080-line, interlaced scanned, 60 fields per second video may be transmitted and then deinterlaced at the receiver.

Another issue is the division of the channel bandwidth between the standard video and video enhancement bits. A 6 MHz channel can currently support approximately 18 megabits per second (Mbps). One migration scheme may involve using 17 Mbps for standard video and 1 Mbps for enhancement data. Another scheme may utilize 16 Mbps for standard video and 2 Mbps for enhancement data. Understanding the effect of the resolution format and the trade-off between bandwidth allocated to the two data types will allow us to determine the best migration schemes to achieve the desired resolution goal in a backward-compatible manner.

2.3 Real-Time Video on the Internet

Sponsors

Advanced Telecommunications Research Program
Lucent Technologies Fellowship

Project Staff

Raynard O. Hinds

The Internet has become sufficient at transmitting data over packet-switched networks. This is because data has no inherent delay constraints and can handle the delay jitter that occurs due to variable queuing delays across the network as well as the excess delay that occurs from retransmission of lost packets. This research has looked at transmitting real-time video over this same network. Real-time video can not tolerate excessive delay. Packets arriving after their scheduled playback point at the receiver are discarded. Video sequences are capable of tolerating loss. Block-based video coders, which rely on motion compensated block prediction for more data compression have been used to code video over networks. With the resulting packet loss that occurs on congested networks, coding mode selection for each macroblock is significant in determining the overall distortion on the decoded video sequence. In the past, a methodology for optimal mode selection in the presence of potential macroblock loss was developed for a restricted set of video coders. For a given channel erasure description and error concealment method, macroblock modes are selected for block-based video coders with zero-motion compensation only to minimize the distortion for a given bit-rate constraint. This algorithm was extended to find the optimal mode selection when motion compensation is allowed. This has led to a more efficient robust coder in the presence of loss.

2.4 Video Compression with Complete Information

Sponsor

Advanced Telecommunications Research Program

Project Staff

David M. Baylon

In many video compression applications, one of the main goals is to deliver the best possible video quality through a bandwidth-limited channel or a capacity-limited storage medium. The video data being compressed often originates from a live video feed, a pre-recorded video source, or alternating segments of both. For a live feed, video processing and compression are essentially causal since the video encoder does not have access to future video data and since real-time constraints dictate that the delay be small and buffer sizes be limited. On the other hand, for a pre-recorded video source, video processing and compression need not be causal since at any instant the video encoder can have access to future video information and can be controlled in a manner which anticipates future video events in order to improve quality.

Traditional video compression algorithms have focused primarily on causal processing of the video data. However, because there exists a wealth of pre-recorded video sources including movies, documentaries and other programs which are recorded on film, there is a great motivation and potential for using noncausal processing for video. This research proposes using pre-processing for pre-recorded video programs to obtain useful information for compression. By exploiting information obtained about a program prior to compression, better video quality can be achieved. In particular, as an example of how noncausal processing of a video program can be used to improve performance over causal processing, a heuristic iterative algorithm is developed for a buffer-constrained quantization problem to minimize the number of quantization changes in MPEG-2 intraframe video. Simulations have shown a significant improvement using noncausal processing.

2.5 Generalized Multi-Dimensional Rate-Distortion Based Bit Rate Control

Sponsors

Advanced Telecommunications Research Program
Charles Stark Draper Laboratory

Project Staff

Eric C. Reed

The bit rate produced at the output of a video encoder is inherently variable. In a communication scenario, the bits produced by the encoder are transmitted through a channel at some fixed or variable bit rate. Since the channel bit rate is different from the bit rate produced by the encoder, a buffer is needed to match the two rates. The buffer size is proportional to a delay constraint that depends on the application. The goal of a bit rate controller is to control the level of the buffer.

Most bit rate control schemes adjust a quantizer parameter to control the buffer level. In the source coding community, a great deal of focus has been placed on how to choose the quantizers to maximize video quality under a buffer constraint. At very low bit rates, the buffer constraint often leads to poor video quality. This thesis focuses on the development of generalized multi-dimensional (M-D) rate-distortion based bit rate control algorithms. In this scheme, rather than adjusting only the quantizer parameter, a vector of encoding parameters is jointly adjusted to control the bit rate (i.e., temporal subsampling factor). We view the problem of video coding as choosing the appropriate operating point, in time, from a set defined on a M-D grid, under a buffer constraint. Since many operating points achieve approximately the same rate, both rate and distortion are considered. The advantage of this scheme is that each parameter (i.e., quantizer) can change at a much slower pace in comparison to a scheme that adjusts only one of the parameters.

As a case study, we experiment with the compression of underwater images, taken from an unmanned undersea vehicle (UUV) down to below 10 kilobits per second (Kbps). The goal is to transmit these images to the mother ship at the ocean surface.

2.6 Compression of Underwater Video Sequences using Quantizer/ Position Dependent-Encoding

Sponsors

Advanced Telecommunications Research Program
Charles S. Draper Laboratory

Project Staff

Theresa Huang

Transmission of underwater video sequences obtained by an unmanned underwater vehicle to a mother ship requires a transmission channel with a bit rate capacity on the order of megabits per second. The transmission channel available has a bit rate capacity of 10 Kbps. In order to obtain such low bit rates, the video sequence is first preprocessed. Motion estimation is used to reduce the temporal redundancies, and an 8×8 block discrete cosine transform (DCT) is applied to minimize the spatial redundancies. Each 8×8 block of DCT coefficients is then encoded into a bit stream. Typically, one standard codebook is used to assign codewords to each event that occurs. Position-dependent encoding (PDE) further reduces the bit rate by exploiting the varying statistical properties of the runlength and the amplitude as a function of the starting position. Coefficients in the lower frequencies are most likely to have a large amplitude and a short runlength, whereas coefficients in the higher frequencies are most likely to have zero amplitude and a long runlength. Further compression is achieved by using quantizer/position-dependent encoding. Not only do the statistics vary with position, they also change as a function of the quantizer. When a very fine quantizer is used, very little data is thrown away. However, when a coarse quantizer is used, most of the DCT coefficients become zero. As a result, quantizer/position-dependent codebooks significantly re-duces the bit rate.

2.7 Speech Enhancement

Sponsors

GEM Fellowship
Innovative Development and Enterprise Advancements Program
U.S. Federal Bureau of Investigation

Project Staff

Rubén E. Galarza

Enhancement of speech degraded by additive noise has been an active area of research for the past 20 years. Many algorithms have been developed to try and solve this problem, each with varying degrees of success. This project focuses on one such technique, which achieves enhancement by segmenting speech into (nearly) stationary regions in the time-frequency plane.

The speech signal is divided into stationary regions through filtering and adaptive windowing. A filter bank is used to separate the signal into frequency channels. Then, an adaptive length window is applied to each channel. The length of the window is determined by a similarity measure based on the cross-correlation of spectra in adjacent time segments. This scheme allows the windowing stage to adapt following specific speech characteristics at different regions in frequency. After the stationary regions have been selected, each of them is enhanced using linear prediction modeling and Wiener filtering. The parameters of the linear prediction model and Wiener filter are adjusted according to the region's estimated signal to noise ratio (SNR).

Although preliminary testing using this system showed promising results, a simpler implementation with higher quality output was desired. To this end, several stages of the original system were modified. For example, the adaptive windowing stage was changed to update parameters according to SNR. Also, the enhancement stage was modified so that low SNR regions at higher frequencies are less affected by Wiener filtering than high SNR regions at lower frequencies. These adjustments simplified the algorithm significantly while moderately improving speech intelligibility. More testing is needed to increase intelligibility in the enhanced speech.