Chapter 2. Advanced Telecommunications and Signal Processing Program

Academic and Research Staff

Professor Jae S. Lim

Visiting Scientists and Research Affiliates

Dr. Hae-Mook Jung

Graduate Students

John G. Apostolopoulos, David M. Baylon, Shiufun Cheung, Raynard O. Hinds, Kyle K. Iwai, Peter A. Monta, Julien J. Nicolas, Alexander Pfajfer, Eric C. Reed, Lon E. Sunshine, Carl Taniguchi, Chang Dong Yoo

Technical and Support Staff

Cindy LeBlanc, Denise M. Rossetti

2.1 Introduction

The present television system was designed nearly 40 years ago. Since then, there have been significant developments in technology which are highly relevant to the television industries. For example, advances in very large scale integration (VLSI) technology and signal processing make it feasible to incorporate frame-store memory and sophisticated signal processing capabilities into a television receiver at a reasonable cost. To exploit this new technology in developing future television systems, Japan and Europe have established large laboratories, funded by government or industry-wide consortia. The need for this type of organization in the United States was considered necessary for the broadcasting and equipment manufacturing industries. In 1983 the advanced television research program (ATRP) was established at MIT by a consortium of U.S. companies.

The major objectives of the ATRP are:

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

Research areas of the program include the design of a receiver-compatible advanced television (ATV) system and digital ATV system, as well as the development of transcoding methods. Significant advances have already been made in some of these research areas. A digital ATV system was designed and tested in the fall of 1992 by the FCC for its possible adoption as the U.S. HDTV standard for terrestrial broadcasting. Some elements of this system are likely to be included in the U.S. HDTV standard.

In addition to research on advanced television systems, our research program also includes research on speech processing. Current research topics include the development of a new speech model and algorithms to enhance speech degraded by background noise.

2.2 ATRP Facilities

The ATRP facilities are currently based on a network of eight SUN workstations and three DecStation 5000 workstations. There is approximately 14.0 GB of disk space, distributed among the various machines. Attached to one of the SUN workstations is a VTE display system with 256 MB of RAM. This display system is capable of driving the Sun-4 monitors, or a 29-inch Conrac monitor, at rates up to 60 frames/sec. In addition to displaying high-resolution real-time sequences, the ATRP facilities include a Metheus frame buffer which drives a

Sony 2K×2K monitor. For hard copy output, the lab uses a Kodak XL7700 thermal imaging printer which can produce 2K×2K color or black and white images on 11×11 inch photographic paper.

Other peripherals include tape drives (Exabyte 8 mm and AMPEX DST600), a 16-bit digital audio interface with two channels and sampling rates up to 48 kHz per channel, and an audio workstation with power amplifier, speakers, CD player, tape deck, etc. In addition, the lab has a 650 MB optical disk drive, a CD-ROM drive, two laser printers, and one color printer. The ATRP facilities also include a PowerMacintosh, two Mac II computers, and an Apple LaserWriter for preparing presentations.

A fast network (FDDI) is under consideration to augment the current 10 Mbps Ethernet. The new network would enable much faster data transfer to display devices and would support large NFS transfers more easily.

2.3 Signal Representations for Very-low-bit-rate Video Compression

Sponsors

AT&T Fellowship
Advanced Telecommunications Research Program

Project Staff

John G. Apostolopoulos

Video, which plays an important role in many applications today, is expected to become even more important in the near future with the advent of multimedia and personal communication devices. The large amount of data contained in a video signal, together with the limited transmission capacity in many applications, requires the compression of a video signal. A number of video compression algorithms have been developed for different applications from video-phone to high-definition television. These algorithms perform reasonably well at respective bit rates of 64 kb/s to tens of Mb/s. However, many applications, particularly those associated with portable or wireless devices, will probably be required in the near future to operate at considerably lower bit rates, possibly as low as 10 kb/s. The video compression methodologies developed to date cannot be applied at such low bit rates. The goal of this research is to create efficient signal representations which can lead to acceptable video quality at extremely low bit rates.

Conventional video compression algorithms may be described as block-based coding schemes; they partition each frame into square blocks and then independently process each block. Examples of these compression techniques include block-based temporal motion-compensated prediction and spatial block discrete-cosine transformation. Block-based processing is the basis for virtually all video compression systems today because it is a simple and practical approach for achieving acceptable video quality at required bit rates. However, block-based coding schemes can not effectively represent a video signal at very low bit rates because the source model is extremely limited: Block-based schemes inherently assume a source model of (translational) moving square blocks, but a typical video scene is not composed of translated square blocks. In effect, block-based schemes impose an artificial structure on the video signal before encoding, instead of exploiting the structure inherent to a particular video scene.

The goal of this research is to develop signal representations that better match the structure that exists within a video scene. By identifying and efficiently representing this structure, it may be possible to produce acceptable video quality at very low bit rates. For example, since real scenes contain objects, a promising source model is one with two- or three-dimensional moving objects. This approach may provide a much closer match to the structure in a video scene than the block-based schemes mentioned above. Three fundamental issues must be addressed for the success of this approach: (1) appropriate segmentation of the video scene into objects, (2) encoding the segmentation information, and (3) encoding the object interiors. In regard to the third issue, significant statistical dependencies exist in regions belonging to each object and must be exploited. Conventional approaches to encoding arbitrarily shaped regions are typically simple extensions of the block-based approaches, and hence suffer from inefficiencies. A number of novel methods for efficiently representing the interior regions have been developed.

2.3.1 Publication

Apostolopoulos, J., A. Pfajfer, H.M. Jung, and J.S. Lim. "Position-Dependent Encoding." *Proceedings of ICASSP* V: 573-576 (1994).

2.4 Constant Quality Video Coding

Sponsors

INTEL Fellowship
Advanced Telecommunications Research Program

Project Staff

David M. Baylon

Traditional video compression algorithms for fixed bandwidth systems typically operate at fixed targeted compression ratios. A problem with fixed bit rate systems is that video quality can vary greatly within an image and across an image sequence. In regions of low image complexity, quality is high, whereas in regions of high image complexity, quality can be low due to coarse quantization.

Recently, with the development of asynchronous transfer mode (ATM) switching for broadband networks, there has been increasing interest in variable rate video coding. Asynchronous channels can use bandwidth efficiently and provide a time varying bit rate well suited for the nonstationary characteristics of video. By allowing the compression ratio to vary with scene contents and complexity, constant quality video can be delivered.

This research focuses on methods for encoding high-definition video to yield constant high quality video while minimizing the average bit rate. Simple mean square error approaches are insufficient for measuring video quality, therefore distortion functions which incorporate visual models are being investigated. Adaptive quantization schemes to maintain constant quality are being studied also. This study includes development of a spatially and temporally adaptive weighting matrix for frequency coefficients based upon a visual criterion. For example, high-frequency coefficients corresponding to edges can be more visually important than those corresponding to texture. By studying how to vary bit allocation with scene contents to maintain a specified level of quality, a statistical characterization of the variable bit rate video coding can be obtained.

2.5 Signal Representations in Audio Compression

Sponsor

Advanced Telecommunications Research Program

Project Staff

Shiufun Cheung

The demand for high-fidelity audio in transmission systems such as digital audio broadcast (DAB) and high-definition television (HDTV) and in commercial products such as MiniDisc and Digital Compact Cassette has generated considerable interest in audio compression schemes. The common objective is to achieve high quality at a rate significantly smaller than the 16 bits/sample used in current CD and DAT systems. We have been considering applications to HDTV; an earlier implementation, the MIT Audio Coder (MIT-AC), is one of the systems that was considered for inclusion in the U.S. HDTV standard. In this research, we seek to build upon our previous efforts by studying an important aspect of audio coder design: the selection of an appropriate and efficient acoustic signal representation.

In conventional audio coders, short-time spectral decomposition serves to recast the audio signal in a representation that is not only amenable to perceptual modeling but also conducive to deriving transform coding gains. This is commonly achieved by a multirate filter bank, or, equivalently, a lapped transform.

In the first stage of this research, we replace our original uniform multirate filter bank with a nonuniform one. Filter banks with uniform subbands, while conceptually simpler and easier to implement, force an undesirable tradeoff between time and frequency resolution. For example, if the analysis bandwidth is narrow enough to resolve the critical bands at low frequencies, poor temporal resolution can result in temporal artifacts such as the "pre-echo" effect. A nonuniform filter bank, on the other hand, allows enough flexibility to simultaneously satisfy the requirements of good pre-echo control and frequency resolution consistent with critical-band analysis. In our prototype implementation, we use a hierarchical structure based on a cascade of M-band perfect-reconstruction cosine-modulated filter banks. Testing the new representation has shown improvement over a uniform filter bank scheme. While pre-distortion still exists in transient signals, listening sessions show that pre-echo is inaudible with the new filter bank.

In the second stage of this research, we are studying the incorporation of biorthogonality into Malvar's lapped transforms. Lapped transforms are popular implementations of cosine-modulated filter banks. They form the major building blocks in both our uniform and nonuniform signal decomposition schemes. Conventional lapped transforms are designed to be orthogonal filter banks in which the analysis and synthesis filters are identical. By allowing the analysis and synthesis filters to differ, we create a biorthogonal transform with additional degrees of freedom over the orthogonal case. This increased flexibility is very important for achieving a spectral analysis that approximates the critical bands of the human ear well enough for subsequent perceptual modeling.

2.5.1 Publications

Cheung, S., and J.S. Lim. "Incorporation of Biorthogonality into Lapped Transforms for Audio Compression." *Proceedings of ICASSP.* Forthcoming.

Monta, P.A., and S. Cheung. "Low Rate Audio Coder with Hierarchical Filter Banks and Lattice Vector Quantization." *Proceedings of ICASSP* II: 209-212 (1994).

2.6 Pre-Echo Detection and Reduction

Sponsor

Advanced Telecommunications Research Program

Project Staff

Kyle K. Iwai

In recent years, there has been an increasing interest in data compression for storage and data transmission. In the field of audio processing, various varieties of transform coders have successfully demonstrated reduced bit rates while maintaining high audio quality. However, there are certain coding artifacts which are associated with transform coding. The pre-echo is one such artifact. Pre-echos typically occur when a sharp attack is preceded by silence. Quantization noise added by the coding process is normally hidden within the signal. When the coder is stationary over the window length, an assumption breaks down in a transient situation. The noise is unmasked in the silence preceding the attack, creating an audible artifact called a pre-echo.

If the length of the noise can be shortened to about 5 ms, psycho-acoustic experiments tell us that the noise will not be audible. Using a shorter window length shortens the length of the pre-echo. However, shorter windows also have poorer frequency selectivity and poorer coder efficiency. One solution is to use shorter windows only when there is a quiet region followed by a sharp attack.

In order to use adaptive window length selection, a detector had to be designed. A simple detector was implemented which compares the variance within two adjacent sections of audio. In a transient situation, the variance suddenly increases from one section to the next. The coder then uses short windows to reduce the length of the pre-echo, rendering the artifact inaudible.

2.7 Video Source Coding for High-Definition Television

Sponsor

Advanced Telecommunications Research Program

Project Staff

Peter A. Monta

Efficient source coding is the technology for high-definition television (HDTV) which will enable broadcasting over the relatively narrow channels (e.g., terrestrial broadcast and cable) envisioned for the new service. Coding rates are on the order of 0.3 bits/sample, and high quality is a requirement. This work focuses on new source coding techniques for video relating to representation of motion-compensated prediction errors, quantization and entropy coding, and other system issues.

Conventional coders represent video with the use of block transforms with small support (typically 8x8 pixels). Such independent blocks result in a simple scheme for switching a predictor from a motion-compensated block to a purely spatial block; this is necessary to prevent the coder from wasting capacity in some situations.

Subband coders of the multiresolution or wavelet type—with their more desirable localization properties, lack of "blocking" artifacts, and better match to motion-compensated prediction errors—complicate this process of switching predictors, since the blocks now overlap. A novel predictive coding scheme is proposed in which subband coders can combine the benefits of good representation and flexible adaptive prediction.

Source-adaptive coding is a way for HDTV systems to support a more general imaging model than conventional television. With a source coder that can adapt to different spatial resolutions, frame rates, and coding rates, the system may then make tradeoffs among the various imagery types (for example, 60 frames/s video, 24 frames/s film, highly detailed still images, etc.). In general, this is an effort to make HDTV more of an image transport system rather than a least-common-denominator format to which all sources must either adhere or be adjusted to fit. These techniques are also applicable to NTSC to some extent; one result is an algorithm for improved chrominance separation for the case of "3–2" NTSC, that is, NTSC upsampled from film.

2.8 Transmission of HDTV Signals in a Terrestrial Broadcast Environment

Sponsor

Advanced Telecommunciations Research Program

Project Staff

Julien J. Nicolas

High-definition television (HDTV) systems currently being developed for broadcast applications require 15-20 Mbps to yield good quality images for roughly twice the horizontal and vertical resolutions of the current NTSC standard. Efficient transmission techniques must be found in order to deliver this signal to a maximum number of receivers while respecting the limitations stipulated by the FCC for over-the-air transmission. This research focuses on the principles that should guide the design of such transmission systems.

The major constraints related to the transmission of broadcast HDTV include (1) a bandwidth limitation (6 MHz, identical to NTSC), (2) a requirement for simultaneous transmission of both NTSC and HDTV signals on two different channels (Simulcast approach), and (3) a tight control of the interference effects between NTSC and HDTV, particularly when the signals are sharing the same frequency bands. Other considerations include complexity and cost issues of the receivers and degradation of the signal as a function of range.

A number of ideas are currently being studied; most systems proposed to date use some form of forward error-correction to combat channel noise and interference from other signals. The overhead data reserved for the error-correction schemes represents up to 30 percent of the total data, it is therefore well worth trying to optimize these schemes. Current work is focusing on the use of combined modulation/coding schemes capable of exploiting the specific features of the broadcast channel and the interference signals. Other areas of interest include the use of combined source/channel coding schemes for HDTV applications and multiresolution coded modulation schemes.

2.8.1 Publication

Nicolas, J., and J.S. Lim. "On the Performance of Multicarrier Modulation in a Broadcast Multipath Environment." *Proceedings of ICASSP* III: 245-248 (1994).

2.9 Position-Dependent Encoding

Sponsor

Advanced Telcommunications Research Program

Project Staff

Alexander Pfajfer

In typical video compression algorithms, the DCT is applied to the video, and the resulting DCT coefficients are quantized and encoded for transmission and storage. Some of the DCT coefficients are set to zero. Efficient encoding of the DCT coefficients is usually achieved by encoding the location and amplitude of the nonzero coefficients. Since in typical MC-DCT compression algorithms, up to 90 percent of the available bit rate is used to encode the location and amplitude of the nonzero quantized DCT coefficients, efficient encoding of the location and amplitude information is extremely important for high quality compression.

A novel approach to encoding of the location and amplitude information, position-dependent encoding, is being examined. Position-dependent runlength encoding and position-dependent encoding of the amplitudes attempts to exploit the inherent differences in statistical properties of the runlengths and amplitudes as a function of their position. This novel method is being compared to the classical, separate, single-codebook encoding of the runlength and amplitude, as well as to the joint runlength and amplitude encoding.

2.9.1 Publication

Apostolopoulos, J., A. Pfajfer, H.M. Jung, and J.S. Lim. "Position-Dependent Encoding." *Proceedings of ICASSP* V: 573-576 (1994).

2.10 MPEG Compression

Sponsors

U.S. Navy - Office of Naval Research
NDSEG Graduate Fellowship
Advanced Telecommunications Research Program

Project Staff

Eric C. Reed

In a typical MC-DCT compression algorithm, almost 90 percent of the available bit rate is used to encode the location and amplitude of the non zero quantized DCT coefficients. Therefore efficient encoding of the location and amplitude information is extremely important. One novel approach to encoding the location and amplitude information of the nonzero coefficients is position-dependent encoding. Position-dependent encoding, in contrast to single-codebook encoding, exploits the inherent differences in statistical properties of the runlengths and amplitudes as a function of position.

Position-dependent encoding has been investigated as an extension to separate encoding of the runlengths and amplitudes and has proven to provide a substantial reduction in the overall bit rate compared to single-codebook methods. However, MPEG compression does not allow separate encoding of the runlengths and amplitudes. Therefore, this research involves developing a position-dependent extension to encode the runlengths and amplitudes jointly as a single event. Rather than having two separate codebooks for the runlengths and amplitudes, one two-dimensional codebook will be utilized. This method will be compared to conventional approaches as well as the position-dependent encoding approach using separate codebooks.

2.11 HDTV Transmission Format Conversion and the HDTV Migration Path

Sponsor

Advanced Telecommunications Research Program

Project Staff

Lon E. Sunshine

The current proposal for terrestrial HDTV broadcasting allows for several possible transmission formats. Because production and display formats may differ, it will be necessary to convert between formats effectively. Since HDTV will presumably move toward progressive display systems, de-interlacing non-progressive source material will be a key process. This research will consider topics relating to conversion among the six formats being proposed for the U.S. HDTV standard.

As HDTV evolves, it is probable that more transmission formats will be accepted. Furthermore, additional bandwidth may be allocated for some channels (terrestrial and/or cable). This research will consider the issues related to the migration of HDTV to higher resolutions. Backward compatibility and image compression and coding issues will be addressed.

2.12 Removing Degradations in Image Sequences

Sponsor

Advanced Telecommunications Research Program

Project Staff

Carl Taniguchi

The development of two-dimensional noise smoothing algorithms have been an active area of research since the 1960s. Many of the traditional algorithms fail to use the temporal correlation that exists between frames when processing image sequences. However, with the increasing speed of microprocessors and the rising importance of video, three-dimensional algorithms have not only become feasible, but also practical.

Three-dimensional median filters that are sensitive to motion is the first step in using the temporal correlation in images. Existing algorithms of this type effectively reduce to two-dimensional median filters under areas of the image undergoing motion. An improvement in the use of temporal correlation can be obtained by using a motion estimation algorithm before filtering the image with a three-dimensional median filter. Various median filters and motion compensation algorithms will be tested in the presence of noise.

Uniform processing of an image tends to unnecessarily blur areas of the image that are not affected by noise. In this case, a degradation detector may be of practical use.

2.13 Speech Enhancement

Sponsor

Maryland Procurement Office
Contract MDA904-93-C-4180

Project Staff

Chang Dong Yoo

The development of the dual excitation (DE) speech model has led to some interesting insights into the problem of speech enhancement. Based on the ideas of the DE model, a new speech model is being developed. The DE model provides more flexible representation of speech and possesses features which are particularly useful to the problem of speech enhancement. These features, along with a variable length window, are the backbone of the new speech model being developed.

Because the DE model does not place any restrictions on its characterization of speech, the enhancement system based on the DE model performs better than the one based on any of the previous speech models. While a model should be inclusive in its characterization, it should have some restrictions. Specifically, a speech model should pertain to speech. The DE model is somewhat unrestrictive and simple in its characterization of speech. It is solely based on the separation of the voiced and unvoiced components. Whether it makes sense to represent a stop as a voiced and an unvoiced component is only one of many interesting issues which are being investigated. An extension of the DE model which deals with these issues better is currently being studied.

All model-based enhancement methods to date have been formulated on the premise that each segment of speech is stationary for a fixed window length. The performance of these enhancement algorithms depends on the validity of this assumption. We use a variable-length window to capture varying durations of stationarity in the speech. There are several algorithms which adaptively detect changes in auto-regressive model parameters in quasi-stationary signals which have been successfully used in speech recognition. We propose to investigate some of these algorithms. Using a variable length window will allow (1) better and "cleaner" separation of the voiced and unvoiced components, and (2) a greater reduction in the number of characteristic parameters, such as the amplitudes of the voiced components and the LP coefficients of the unvoiced component.

Professor Emeritus William F. Schreiber (Photo by John F. Cook)