Chapter 2. Advanced Telecommunications and Signal Processing Program

Academic and Research Staff

Professor Jae S. Lim

Visiting Scientists and Research Affiliates

M. Carlos Kennedy

Graduate Students

John G. Apostolopoulos, David M. Baylon, Shiufun Cheung, Raynard O. Hinds, Peter A. Monta, 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 50 years ago. Since then, there have been significant developments in technology, which are highly relevant to the television industry. 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, Japan and Europe established large laboratories, funded by government- or industry-wide consortia. The lack of this type of organization in the United States was considered detrimental to the broadcasting and equipment manufacturing industries, so in 1983 a consortium of American companies established the Advanced Television Research Program (ATRP) at MIT.

The major objectives of the ATRP are:

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

In the past, 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 are likely to be included in the U.S. HDTV standard. Specifically, the ATRP group represented MIT 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 advanced television (ATV) standard in the United States. This standard was formally recommended for adoption as the American ATV standard by the FCC's Advisory Committee on Advanced Television Service.

In addition to research on issues related to the design of digital television system, our program currently includes research on telecommunication issues and speech enhancement.

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

Sponsors

Advanced Telecommunications Research Program
AT&T Fellowship

Project Staff

John G. Apostolopoulos

Video plays an important role in many of today's telecommunications applications, and it is expected to gain even greater importance in the near future with the advent of multimedia and personal communication devices. The large raw data rate of a video signal, together with the limited transmission capacity in many applications, necessitates compression of the 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 in the near future, particularly those associated with portable or wireless devices, will most likely be required to operate at considerably lower bit rates, possibly as low as 10 kb/s. The video compression methodologies developed thus far are not applicable at such low bit rates. The goal of this research is to create efficient signal representations which can lead to acceptable video quality at these 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 (1) block-based temporal motion-compensated prediction and (2) 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 to achieve acceptable video quality at the 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. However, a typical video scene is not composed of translated square blocks. In effect, block-based schemes impose an artificial structure on the video signal and then try to encode this structure, as opposed to recognizing the structure inherent to a particular video scene and attempting to exploit it.

The goal of this research is to develop signal representations that are better matches for 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 two- or three-dimensional moving objects. This approach may provide a much closer match to the structure in a video scene than the aforementioned block-based schemes. Three fundamental issues that must be addressed for the success of this approach are (1) appropriate segmentation of the video scene into objects, (2) encoding of the segmentation information, and (3) encoding of 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 encode arbitrarily shaped regions are typically simple extensions of the block-based approaches, and hence suffer from inefficiencies. A number of novel methods to efficiently represent the region interior have been developed.

2.2.1 Publications

Apostolopoulos, J.G., and J.S. Lim. "Representing Arbitrarily-shaped Regions - A Case Study in Overcomplete Representations." Paper presented at the IEEE International Conference on Image Processing, Washington, D.C., October 1995.

Apostolopoulos, J.G., and J.S. Lim. "Coding of Arbitrarily-Shaped Regions." Paper presented at the Conference on Visual Communications and Image Processing, Taipei, Taiwan, May 1995.

2.3 Optimal Transform Coefficient Selection for Images

Sponsors

Advanced Telecommunications Research Program
INTEL Fellowship

Project Staff

David M. Baylon

In many image coding schemes, compression is achieved by perceptual quantization and thresholding of data. In transform coding, thresholding of high-frequency, low-energy coefficients can lead to substantial compression while maintaining high image quality. However, in applications where it is desirable to optimize performance while keeping the rate or image quality constant, it becomes increasingly important to select coefficients in a more optimal fashion. This research focuses on optimal transform coefficient selection in a rate-distortion framework using mean square error distortion.

Traditional threshold schemes which select coefficients based upon magnitude are very fast but do not guarantee to generate optimal solutions since consideration is not given to the bit rate requirement of encoding the selected coefficients. Previous convex hull methods are fast and typically yield optimal solutions when performed globally over the image but not when performed locally.

In this research, an algorithm is proposed for methodically deriving rate-distortion points by iteratively generating a set of convex hulls, from which a composite operational rate-distortion curve is generated. Although this "composite shell" method can be used for generating both globally and locally optimal rate-distortion points, the complexity of this approach is high. Therefore, a fast suboptimal approach is also proposed which is based upon a modified version of threshold selection. Simulations performed on some typical images using finely quantized DCT coefficients and separate coding of amplitudes and runlengths show that very good rate-distortion performance can be obtained using this fast algorithm while maintaining decoder-compatibility.

Current research is aimed at further improving performance by designing the analysis and synthesis filters to compensate for the distortion introduced by quantization and selection.

2.4 Biorthogonality in Lapped Transforms: A Study in Audio Compression

Sponsor

Advanced Telecommunications Research Program

Project Staff

Shiufun Cheung

The demand for high-quality audio in transmission systems such as digital audio broadcast (DAB) and high-definition television (HDTV), as well as commercial products such as the MiniDisc (MD) and the digital compact cassette (DCC), 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 compact disc (CD) and digital audio tape (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 build upon our previous efforts by studying one important aspect of audio coder design: the short-time spectral decomposition.

In conventional audio coders, the 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 gain. This decomposition is commonly achieved by a multirate filter bank, or equivalently, a lapped transform.

Towards the goal of improving the performance of audio compression schemes, we have formulated a biorthogonal cosine-modulated filter bank which is a generalization of Malvar's extended lapped transform (ELT). The ELT, a popular implementation of cosine-modulated filter banks, is of particular interest because it forms the major building block of signal decomposition schemes in many audio coders.

Conventional lapped transforms are designed to be orthogonal filter banks in which the analysis and synthesis filters are identical. Allowing the analysis and synthesis filters to differ leads to a biorthogonal transform which has more degrees of design freedom than its orthogonal counterpart. We have proven that the incorporation of biorthogonality into an M-channel ELT yields an increase of $M/2$ degrees of freedom. This additional flexibility allows the design of synthesis filter banks with improved sidelobe behavior which should be beneficial to audio coder performance.

2.4.1 Publication

Cheung, S., and J.S. Lim. "Incorporation of Biorthogonality into Lapped Transforms for Audio Compression." *Proc. ICASSP* 5: 3079-3082 (1995).

2.5 Multistage Video Compression

Sponsors

Advanced Telecommunications Research Program
AT&T Fellowship

Project Staff

Raynard O. Hinds

The television production and broadcasting process can be divided into three broad categories: production, post-production, and broadcasting. During the production stage, the uncompressed television signal is acquired at a bit rate of 1.2 Gbits/sec. The video data will require further processing and editing before the final signal is transmitted to the consumer. Editing is done during the post-production stage where the television signal is manipulated and processed to produce special effects. During this stage, it is desirable to have a compressed representation of the signal to reduce the amount of storage required. The television signal must also be compressed for the final stage which consists of terrestrial broadcast to the consumer. The final two stages of the process require compressed representations. The requirements for the compression algorithm at each stage differ and may require the concatenation of separate algorithms. This research will consider a multistage compression algorithm to determine what problems may exist in its application to television broadcasting.

Post-production of video signals includes rearrangement of images, insertion of text and graphics, as well as temporal filtering for motion effects. In many cases, it is necessary to be able to access and manipulate any frame and insert it into the coded bit stream. This requires random access of frames in the compressed bit-stream. During this stage, there is further processing of the video signal, and it is necessary to keep the image at a quality sufficient for post-production.

For terrestrial broadcast, the television signal must achieve a bit rate of approximately 20 Mbits/sec to fit within the allocated 6 MHz channel. Thus, a high compression ratio is needed. For HDTV transmission a standardization process is taking place which will determine the algorithm applied at this stage to meet the requirements. The image representation at this stage will be fixed. The goal of this research is to find the best representation for the television signal at the first compression stage which meets the specified requirements above to deliver an acceptable image quality to the viewer.

2.5.1 Publications

Hinds, R.O., and T.N. Pappas. "An Adaptive Clustering Algorithm for Segmentation of Video Sequences." *Proc. ICASSP* 4: 2427-2430 (1995).

Pappas, T.N., and R.O. Hinds. "On Video and Audio Data Integration for Conferencing." *Proc. SPIE* 2411: 120-127 (1995).

2.6 Video Source Coding for High-Definition Television

Sponsor

Advanced Telecommunications Research Program

Project Staff

Peter A. Monta

Efficient source coding is the enabling technology for high-definition television over the relatively narrow channels envisioned for the new service (e.g., terrestrial broadcast and cable). 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 8 x 8 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 have more desirable localization properties, lack "blocking" artifacts, and match better to motion-compensated prediction errors. Therefore, they complicate the 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 an image transport system rather than a least-common-denominator format to which all sources either must adhere or be modified 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.

2.7 Improvement of MPEG by Position-dependent Encoding

Sponsors

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

Project Staff

Eric 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 non-zero 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.

2.8 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 in an effective way. A key to this process is the de-interlacing process. Since HDTV will presumably move toward progressive display systems, it will be necessary to de-interlace nonprogressive source material. The 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 allowed. 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.

2.9 Removing Degradations in Image Sequences

Sponsor

Advanced Telecommunications Research Program

Project Staff

Carl Taniguchi

The development of two-dimensional noise smoothing algorithms had 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.

Developing 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.

2.10 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 this model performs better than systems 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 just 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. To improve the performance of the enhancement algorithm, this assumption of stationarity must be assured. In order to do so, a variable-length window should be used 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. The benefit from using a variable length window is two-fold: (1) it will allow better and "cleaner" separation of the voiced and unvoiced components, and (2) it will allow for 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.